Methods and strategies for improving the novel view synthesis quality of neural radiation field

Shun Fang[1], Ming Cui[1,], Xing Feng[1], Yanna Lv[1]

Abstract

Neural Radiation Field (NeRF) technology can learn a 3D implicit model of a scene from 2D images and synthesize realistic novel view images. This technology has received widespread attention from the industry and has good application prospects. In response to the problem that the rendering quality of NeRF images needs to be improved, many researchers have proposed various methods to improve the rendering quality in the past three years. The latest relevant papers are classified and reviewed, the technical principles behind quality improvement are analyzed, and the future evolution direction of quality improvement methods is discussed. This study can help researchers quickly understand the current state and evolutionary context of technology in this field, which is helpful in inspiring the development of more efficient algorithms and promoting the application of NeRF technology in related fields.

Keywords: NeRF, 3D reconstruction, New view synthesis, Volume rendering

1. Introduction

New view synthesis and 3D modeling are key research areas in computer vision. In recent years, researchers have increasingly focused on the NeRF[1] in this regard. NeRF provides an accurate and simple method to represent 3D scenes useing an implicit function based on MLPand, and has achieved satisfactory rendering quality in 3D reconstruction tasks. Current efforts aim to extend the original NeRF to different situations, such as scene synthesis[2, 3], dynamic scenes[4, 5], large scene reconstruction[6, 7] or rapid convergence[8, 9], among others.

Since NeRF was published in 2020, the NeRF paper has been cited more than thousands of times in the past three years. In addition, researchers have made numerous improvements to the NeRF technology. Some work have focused on optimizing the rendering speed of NeRF[10, 11], while others have explored different application scenarios[12, 13]. Futhermore, there have been efforts to extended NeRF for scene inpainting[14, 15], texture synthesis[16], handing complex scenes[17], and addressing more challenging problems.

Email addresses: `fangshun@ieee.org` (Shun Fang), `cuiming@lumverse.com` (Ming Cui), `fengxing@lumverse.com` (Xing Feng), `lvyanna@lumverse.com` (Yanna Lv)

A significant drawback of NeRF is that the quality of image rendering needs to be improved. For example, after removing unnecessary objects from an image, the 3D scene repair fails to maintain the appearance consistency and geometric coherence across multiple views, resulting in low-quality synthesized images. It also becomes challenging to directly model subtle mesoscopic structures. Training on sparse input views can lead to overfitting and erroneous scene depth estimation, which causes artifacts in rendered new views. Additionally, NeRF's inherent multi-layer perceptron struggles to learn high-frequency details, leading to blurriness issues in high-resolution renderings. When there are reflective and refractive objects, NeRF can lead to blurry or distorted renderings. The rendering quality issues have hindered the practical deployment of NeRF technology. Fortunately, over the past three years, researchers have made significant efforts to develop an increasing number of techniques, methods, and models to enhance NeRF's rendering quality. As a result, the combination of NeRF with other technologies has achieved excellent results in variety application scenarios.

This article provides a comprehensive review of the latest research on various techniques to enhance the rendering quality of NeRF. It not only introduces variety a lot of methods to improve rendering quality, but also analyzes and summarizes the characteristics and limitations of these methods. Then we analyze the possibility of combining distinct technologies to achieve improved quality effects, which will help stimulate the generation of more efficient algorithms. The ultimate goal is to advance the application of NeRF technology in new view synthesis and other fields.

2. Principle of neural radiation field

For a given three-dimensional scene, the appearance of any location depends on the specific location and viewing angle. The colors displayed by the scene are related to lighting conditions, causing the colors to change when the same location is viewed from different angles. NeRF represents a three-dimensional scene as a radiation field approximated by a neural network. The radiance field describes the volumetric density and color of each point in the scene for every viewing direction, and represents the static scene as a continuous five-dimensional vector function, expressed as follows.

$$(r, g, b, \sigma) = F_\Theta(x, y, z, \theta, \varphi) \tag{1}$$

(x, y, z) represents the three-dimensional coordinates in the scene space, $\boldsymbol{d} = (\theta, \phi)$ represents the viewing direction of the input image, $c = (r, g, b)$ represents the color emitted from this position in the d direction. θ represents volume density, which is similar to differentiable opacity and represents the amount of radiation accumulated when light passes through (x, y, z). F_Θ means that the five-dimensional function is implicitly expressed using one or more MLP neural networks. By inputting a series of images taken from different perspectives, the output is the color c and volume density σ corresponding to the three-dimensional space position. Based on the scene representation foundation of NeRF, classical volume rendering methods can be employed to render new images from different perspectives. Specifically, for any pixel in the image, N points are sampled along the

ray r of the observation angle. For each sampling point, Θ_i and rgb_i are first calculated based on F_Θ, and then calculated according to the following formula Final color value:

$$\overset{\wedge}{rgb} = \sum_{i=1}^{N} rgb_i w_i \tag{2}$$

$$w_i = T_i \alpha_i \tag{3}$$

$$T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) \tag{4}$$

$$\alpha_i = 1 - \exp\left(-\sigma_i \delta_i\right) \tag{5}$$

δ_i represents the sampling interval on ray r.

In order to train the specific parameters of the MLP corresponding to F_Θ, for a given scene, n images are captured using cameras with different postures. By using the gradient descent method, fitting F_Θ by minimizing the error between the predicted image I_p and the ground truth image I_c, that is $\min \sum_{i=1}^{m} |I_p - I_c|^2$.

Figure 1 shows the flow of the NeRF algorithm. In the literature[1], in order to obtain an output image containing more high-frequency information, the author performed a high-order encoding operation $\gamma(\cdot)$ on the position and perspective parameters of the input MLP.
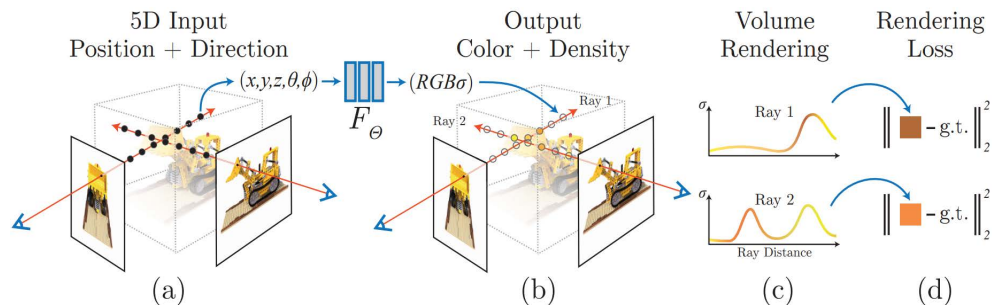

Figure 1: Neural Radiation Field Algorithm Process[1]

The blurry image rendering quality of NeRF is primarily related to the following aspects:
1) Insufficient training data: The NeRF model requires a large amount of high-quality training data to learn the lighting and depth information of the scene. If the training data is limited or of poor quality, the model may not capture rich scene details, resulting in lower image quality.
2) Image resolution: If the resolution of the input image is low, it can affect the quality of the final reconstruction and synthesized images from new viewpoints. However, inputting high-resolution images into the basic NeRF can lead to computational explosions.
3) Scene complexity: The basic NeRF can only handle static images. When dealing with dynamic

scenes, large scenes, unbounded scenes, or complex scenes, the synthesized images will have artifacts such as ghosting, blurring, and inconsistent viewss.

3. An introduction to techniques to improve rendering quality

In view of the huge potential development prospects of NeRF, researchers have proposed a series of technologies to improve the quality of rendered images in the past three years, so that NeRF can be combined with more algorithms and models to handle different tasks. As analyzed in the previous section, the actual rendering quality of NeRF is comprehensively affected by factors such as the amount of training data, the resolution of the input image, and the complexity of the application scenario. This article categorizes and introduces the latest techniques from four aspects: prior guidance, adjusting NeRF model structure, adopting pre-trained models, and high-resolution image processing.

3.1. Prior guidance

In 3D surface reconstruction and new view synthesis, prior guidance refers to the process of assisting relative tasks through known prior knowledge or models to improve the accuracy and stability of task results. The prior information can be object shape, texture, posture, motion, etc. Prior guidance plays an important role in 3D surface reconstruction and new view synthesis, improving the quality and accuracy of the results and shortening the required time and computational resources. Different prior guidance can be selected and combined according to specific application scenarios and requirements. In a series of NeRF papers, many researchers have utilized prior knowledge to achieve rendering in different application scenarios, ultimately achieving higher visual quality.

As is well known, each scene of NeRF requires at least 50-100 view inputs for training in order to synthesize new views from arbitrary viewpoints. However, when the number of input views is not allowed, training will lead to artifacts in the rendered new views. Therefore, introducing prior guidance can solve such problems and improve the rendering quality of the images. ViP-NeRF[18] and DiffusioNeRF[19] both introduce prior knowledge and combine it with NeRF to improve the rendering quality of synthesized images from new viewpoints under sparse input view conditions.

ViP-NeRF computes the visibility priors between views by assuming that the visibility of pixels in different input views can provide a more reliable estimate to supervise the density. The specific implementation steps are as follows: 1) visibility regularization; 2) visibility prior; 3) effective prediction of visibility. The core is in the second step. where given a primary view and an arbitrary secondary view, the objective is to estimate whether each pixel in the main view is also visible in the secondary view through the visibility prior. Visibility is calculated using a plane sweep volume, which does not require any pre-training. By regularizing NeRF with visibility priors, NeRF is successfully trained with few input views. In this paper, the plane scan volume is used to compute the visibility prior, as shown in Figure 2. The plane scan volume is created by warping the different depth planes of the secondary view image to the main view, and compared with the main view

4

image to obtain the error map. It can be found from the figure that the second and third planes match the main view better than the other two planes, resulting in lower errors. The minimum error of all planes is then thresholded to obtain a visibility prior map corresponding to the primary view image. Finally, on the RealEstate-10K data set, the PSNR index of ViP-NeRF is 3-4 points higher than the latest paper in 2, 3, and 4 input views.
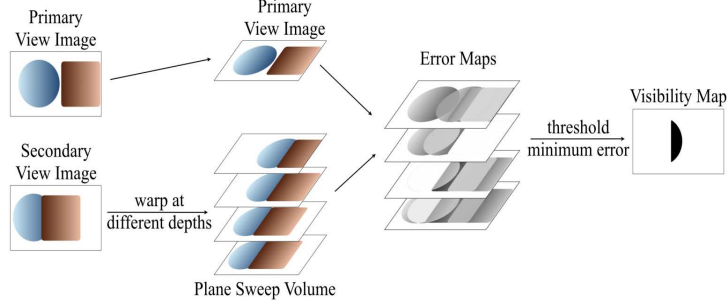

Figure 2: Visibility prior calculation[18]

DiffusioNeRF utilizes a denoising diffusion model to learn priors for scene geometry and color. A denoising diffusion model is trained on RGBD patches from the synthetic Hypersim dataset can be used to predict the logarithmic gradient of the joint probability distribution of color and depth patches. The logarithmic gradient of the priors for RGBD patches helps normalize the geometry and color of the scene. During NeRF training, random RGBD patches are rendered and the estimated gradients of the log-likelihood is backpropagated to the color and density fields. The learned prior achieves better quality in the reconstructed geometry. The PSNR indicators in the 3, 6, and 9 input views of the LLFF dataset are 1-2 points higher. Among the three input views, the PSNR index is nearly 1 point higher than ViP-NeRF.

NoPe-NeRF[20] introduces undistorted monocular depth prior to address the challenging problem of training a NeRF without precomputing camera poses. The prior is generated by correcting the shift parameters and scale during training and can be used to constrain the relative poses between consecutive frames, which is achieved using a novel loss function proposed in the paper. NoPe-NeRF demonstrates superior performance in terms of novel view rendering quality and pose estimation accuracy compared to existing methods.

In terms of prior guidance, NeRFVS[21] proposes an overall prior for neural reconstruction to improve NeRF and realize free view synthesis in indoor scenes. DP-NeRF introduces physical scene priors to handle challenging scenes with blur and exposure[22], due to the sparsity of point cloud representation, utilizing point cloud 3D priors to synthesize high-quality images from colored point clouds is often used for novel indoor scenes[23]. All these approaches have improved the quality of rendered images in NeRF across different tasks.

NeRF was initially applied to new viewpoint synthesis of static images, but as scenes become more complex, NeRF may need to adjust the basic MLP model structure to flexibly handle the task, such as changing the levels and number of MLPs to encode more high-frequency information, or modifying MLP input and output information to improve rendering quality.

MS-NeRF[24] addresses the rendering problem of distortion and blurring caused by reflection or refraction in NeRF. While NeRF satisfies the principle of multi-view consistency for new viewpoint synthesis, there is no consistency in the presence of mirrors in the scene. The paper proposes a multi-space neural radiation field. In order to understand the existence of refractive and reflective objects, the scene is represented by a set of characteristic fields in parallel subspaces. The specific model architecture is shown in Figure **??**. where the output and volumetric rendering module of NeRF are modified to output density and features instead of density and color. Multiple sets of density and features are outputted. Volumetric rendering then generates multiple feature maps. Decoder MLP is applied in each subspace to decode RGB images, and Gate MLP obtains pixel-level weights for the feature maps. The features from each subspace are integrated by summation to obtain the final result. The advantage of MS-NeRF is that it can automatically handle mirror objects in 360° scene rendering. On synthetic data sets, real data sets and RFFR data sets, the PSNR index is 2-4 points higher.

In order to separate the appearance materials and lighting to better relight the high-frequency environment and allow the appearance to be edited, DE-NeRF[25] is proposed to decompose the lighting, geometry and appearance of the input scene. The network architecture is shown in Figure **??**, which is divided into three modules: geometric reconstruction, scene decomposition, and relighting. The geometric reconstruction is to train a signed distance field (SDF) function by inputting a set of 2D images to reconstruct the display grid as the input scene of MLP. Different from NeRF, DE-NeRF decomposes the scene into geometric features and appearance features (including diffuse reflection, roughness, specular tint), combined with the distance h from the sampling point to the grid, is input into four corresponding MLPs to predict diffuse albedo, roughness value, and specular tint. Then integrate the diffuse reflectance with the environment map to get the diffuse color c_d, train a specular lighting MLP F_s to get the specular lighting c_l, and then multiply it with the specular hue to get the specular color c_s, $c_s + c_d$ gets the final color$c$. The final quality is 2 points higher than the PSNR index of the latest paper.

NeRF requires 3D supervision to be realized and does not contain semantic and instance information. The proposal of Nerflets[26] can achieve efficient structure-aware local radiation fields for 3D scene representation with only 2D supervision, which is a novel form of 3D scene representation. Nerflets represent a set of local neural radiance fields, collectively representing a scene, with each Nerflet possessing orientation, spatial position, and range. This allows for panoramic reconstruction within their respective ranges. In contrast to NeRF, Nerflets do not utilize a single large MLP to obtain the color and density of sampled points; instead, the influence weights of each Nerflet are combined. Specifically, for a sampled point, N small MLPs generate N densities and colors, one for each Nerflet, and then the final color and density of the sampled point are obtained through

weighted averaging based on the influence weights. Nerflets can be utilized for 3D reconstruction, novel view synthesis, and direct scene editing. The PSNR metric also shows an improvement of more than one point. The model framework is shown in Figure 3.

There are also NeRFLight[27] and RelightingNeRF[28] that change the underlying model architecture. The former extends the grid-based method by dividing the scene into eight different regions, effectively splitting NeRF into eight separate density fields and color fields. Each region corresponds to an MLP that outputs density and color. In the eight different regions, a shared feature grid is used, allowing features to be symmetrically placed at the center of the scene. This approach avoids the need to increase the number of features required for seamless reconstruction while reducing the number of feature values that need to be stored. The latter is proposed for handling unstructured photographs and introduces a new NeRF approach for free-viewpoint reconstruction. The model architecture is changed by inputting the sampling point position into the first MLP, which outputs SDF values and latent features. From the SDF values, we can obtain normals, shadow cues, and specular cues. By providing sampling point positions, lighting positions, and viewing directions, along with the latent features output by the first MLP, the second MLP produces the final color. This approach results in improved rendering quality.
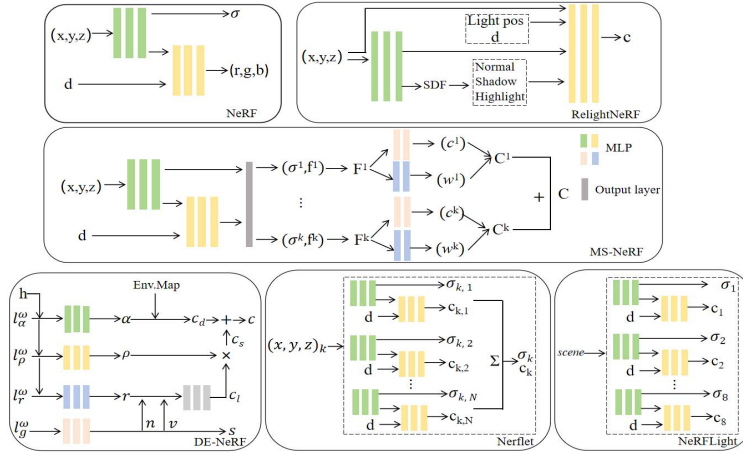

Figure 3: neural network architectures of NeRF, MS-RNeF, DE-NeRF, Nerflets, NeRFLight, and RelightingNeRF

3.3. Use pre-trained models

Another extension of NeRF to improve image rendering quality is to use pre-trained models to optimize the NeRF model for better performance. Latent-NeRF[29] applies score distillation to a latent diffusion model, and then applies the entire process to the compact latent space of a pre-trained autoencoder. The diffusion model denoises the latent encodings of a pre-trained autoencoder. Score distillation, which is commonly used in neural rendering and 3D object generation, is applied to train the model using the diffusion model as the loss function. The specific implementation of Latent-NeRF is shown in Figure 4, where the MLP takes position and direction information

as input and outputs the volume density and four pseudo color channels in a compact latent space. The output here differs from the basic NeRF in that the four pseudo color channels correspond to the four latent features used in the diffusion model. Then, a feature map z is rendered from any viewpoint of the scene, and the resulting image is noised to generate a noisy image. The stable diffusion model is then used along with guidance from text and shape to denoise the image. Finally, the denoising process is backpropagated to the NeRF representation. Latent-NeRF is the first approach that allows coloring of the grid under the guidance of a pre-trained diffusion model, achieving impressive results.
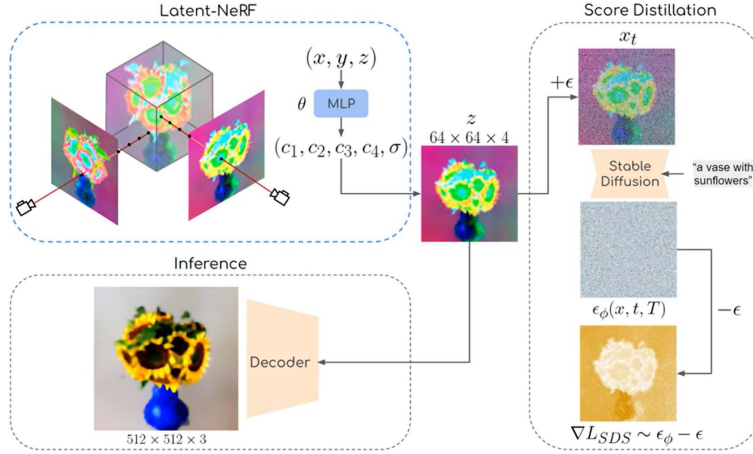

Figure 4: Latent nerf for fractional distillation training[29]

NeRDi[30] is a single-view NeRF synthesis framework without 3D supervision that optimizes NeRF representation by minimizing the diffusion loss on arbitrary view rendering through a pre-trained image diffusion model. The specific model framework is depicted in Figure 5. Given an input image, and a pre-trained visual language model is used to introduce two language guides as the conditioned reflex input of the diffusion model. s_0 represents the image title containing the overall semantics, while s^* is the text embedding extracted from the input image, which captures visual clues. These two language guides help improve the consistency of multi-view content. In addition, the paper also introduces a geometric loss based on estimated depth mapping to regularize the underlying 3D geometry of NeRF. These efforts can synthesize a more realistic new viewpoint image, with PSNR indicators twice as that of high as traditional NeRF, and are also applicable in zero-shot NeRF synthesis.
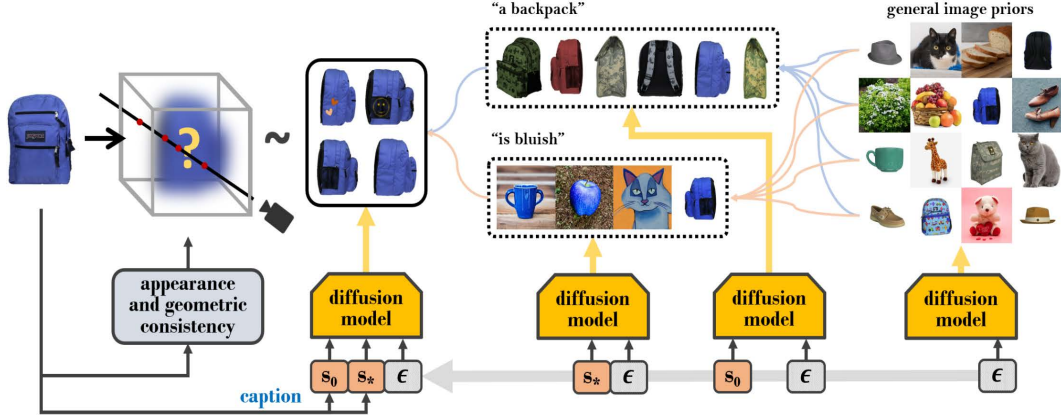
Figure 5: NeRDi model architecture[29]

3.4. High resolution image processing

Although NeRF has achieved remarkable results, there is limited discussion among researchers regarding its limitations in high-resolution image processing. Because NeRF's MLP has a drawback: it struggles to capture high-frequency details, especially as the resolution increases. This leads to challenges such as an explosion of parameters and computational complexity, as well as misalignment of input data. Therefore, this section introduces the latest techniques for high-resolution processing in NeRF.

RefSR-NeRF[31] proposes a reference-guided super-resolution neural radiation field framework, which ultimately achieves high-quality new perspective synthesis by combining NeRF and RefSR-CNN models, and balances memory and speed. The framework process is shown in Figure 6. In the first stage, the sampling strategy of NeRF is modified. While NeRF uses batch ray sampling, RefSR-NeRF adopts patch ray sampling on the high-resolution input image to obtain a low-resolution image. This operation reduces memory storage requirements and improves the speed of NeRF. In the second stage, the obtained low-resolution image and the provided high-resolution guidance image are inputted to RefSR-CNN. This process yields feature maps from the detail-dominant information module and the degradation-dominant information module within RefSR-CNN. These two feature maps are then fused to reconstruct a high-resolution image, resulting in a PSNR metric improvement of nearly one point compared to the state-of-the-art papers.

AligNeRF[32] also discusses the limitations of training NeRF on high-resolution images through two improvements. 1) Combining MLP with convolutional layers to encode more neighborhood information and reduce the number of parameters; 2) Propose new strategies to address the impact of moving objects. The paper introduces a specific alignment-aware strategy that can correct the inaccuracies in alignment, which often lead to performance degradation. By improving the representational capacity of the neural radiance field, this strategy significantly enhances the quality of rendered images. AligNeRF not only improves rendering quality when training high-resolution images, but also does not significantly increase training time. From the PSNR index, we can find

that AligNeRF is better than the best mip-NeRF result, and is more than 3 points higher than NeRF.
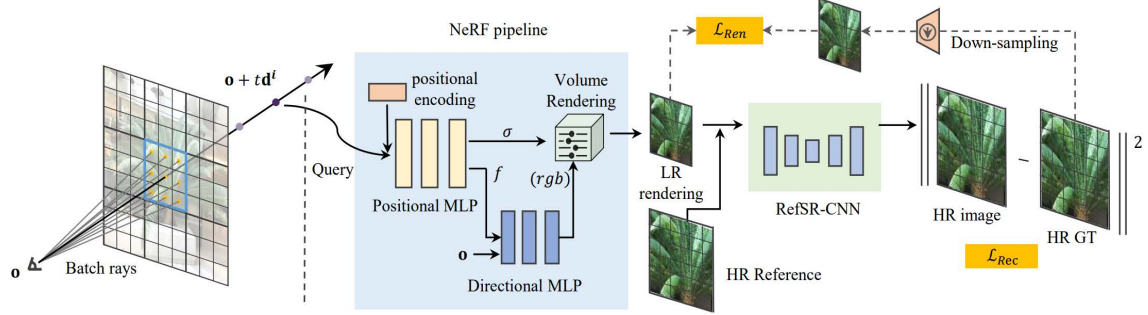

Figure 6: Overview of the RefSR-NeRF framework[31]

4. Summary and Outlook

Since its introduction in 2020, NeRF has provided a new perspective on view synthesis and 3D reconstruction, leading to rapid development and extensive discussions among researchers over the past three years. In the future, NeRF technology will continue to bring significant transformations to new domains.

This article mainly discusses the image quality issues in the new perspectives of NeRF technology for view synthesis and 3D reconstruction. The factors that affect the rendering quality include insufficient training data, low image resolution, and high scene complexity. The article reviews the latest technologies from four aspects: prior guidance, modification of the base NeRF model structure, using pre-trained models, and high-resolution image processing. It introduces the principles and specific implementation methods of each technology for enhancing rendering quality based on NeRF.

NeRF technology is still rapidly evolving and there is a need for further advancements to enhance rendering quality in various real-world scenarios. Looking forward to future technological developments, we should focus on the following research directions.

(1) Applied to large-scale scenes and unbounded scenes Research on view synthesis and 3D reconstruction of large scenes and unbounded scenes has always attracted the attention of researchers. However, due to the complex configuration and dynamic changes of such scenes, some need to provide semantic logic or perform instance segmentation in advance, and others require information such as accurate depth estimation,which can limit the model's ability to express complex content. These types of scenarios could be addressed in the future by applying Normalized Device Coordinates (NDC) or Multi-Sphere Imagery (MSI). Additionally, an unknown pose NeRF model can be used to construct an end-to-end model for large-scale scene modeling.

(2) Accelerating rendering speed In order to improve rendering quality, it is sometimes necessary

to sacrifice rendering speed, such as increasing training time, adding more parameters, and enhancing image resolution. In the future, the goal is to improve rendering image quality without compromising rendering speed.

This can be achieved by first training the MLP model within NeRF using prior guidance to enhance rendering quality, and then distilling the knowledge to train a more compact MLP model. we can also Multi-threaded parallel processing scenarios; or use high-performance chips, because the development of chips and algorithms complement each other. By selecting chips that offer better matching in terms of performance, cost, power consumption, and other factors, the application of the algorithm can be further promoted.

References

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng, Nerf: Representing scenes as neural radiance fields for view synthesis, Communications of the ACM 65 (1) (2021) 99–106.

[2] A. Kundu, K. Genova, X. Yin, A. Fathi, C. Pantofaru, L. J. Guibas, A. Tagliasacchi, F. Dellaert, T. Funkhouser, Panoptic neural fields: A semantic object-aware neural scene representation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12871–12881.

[3] B. Yang, Y. Zhang, Y. Xu, Y. Li, H. Zhou, H. Bao, G. Zhang, Z. Cui, Learning object-compositional neural radiance field for editable scene rendering, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 13779–13788.

[4] T. Li, M. Slavcheva, M. Zollhoefer, S. Green, C. Lassner, C. Kim, T. Schmidt, S. Lovegrove, M. Goesele, R. Newcombe, et al., Neural 3d video synthesis from multi-view video, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5521–5531.

[5] L. Wang, J. Zhang, X. Liu, F. Zhao, Y. Zhang, Y. Zhang, M. Wu, J. Yu, L. Xu, Fourier plenoctrees for dynamic radiance field rendering in real-time, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 13524–13534.

[6] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, P. Hedman, Mip-nerf 360: Unbounded anti-aliased neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5470–5479.

[7] Y. Xiangli, L. Xu, X. Pan, N. Zhao, A. Rao, C. Theobalt, B. Dai, D. Lin, Bungeenerf: Progressive neural radiance field for extreme multi-scale scene rendering, in: European conference on computer vision, Springer, 2022, pp. 106–122.

[8] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, A. Kanazawa, Plenoxels: Radiance fields without neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5501–5510.

[9] C. Sun, M. Sun, H.-T. Chen, Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5459–5469.

[10] T. Müller, A. Evans, C. Schied, A. Keller, Instant neural graphics primitives with a multiresolution hash encoding, ACM Transactions on Graphics (ToG) 41 (4) (2022) 1–15.

[11] T. Hu, S. Liu, Y. Chen, T. Shen, J. Jia, Efficientnerf efficient neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12902–12911.

[12] D. Levy, A. Peleg, N. Pearl, D. Rosenbaum, D. Akkaynak, S. Korman, T. Treibitz, Seathrunerf: Neural radiance fields in scattering media, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 56–65.

[13] A. Ruzzi, X. Shi, X. Wang, G. Li, S. De Mello, H. J. Chang, X. Zhang, O. Hilliges, Gazenerf: 3d-aware gaze redirection with neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 9676–9685.

[14] A. Mirzaei, T. Aumentado-Armstrong, K. G. Derpanis, J. Kelly, M. A. Brubaker, I. Gilitschenski, A. Levinshtein, Spin-nerf: Multiview segmentation and perceptual inpainting with neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 20669–20679.

[15] S. Weder, G. Garcia-Hernando, A. Monszpart, M. Pollefeys, G. J. Brostow, M. Firman, S. Vicente, Removing objects from neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 16528–16538.

[16] Y.-H. Huang, Y.-P. Cao, Y.-K. Lai, Y. Shan, L. Gao, Nerf-texture: Texture synthesis with neural radiance fields, in: ACM SIGGRAPH 2023 Conference Proceedings, 2023, pp. 1–10.

[17] J. Sun, X. Chen, Q. Wang, Z. Li, H. Averbuch-Elor, X. Zhou, N. Snavely, Neural 3d reconstruction in the wild, in: ACM SIGGRAPH 2022 Conference Proceedings, 2022, pp. 1–9.

[18] N. Somraj, R. Soundararajan, Vip-nerf: Visibility prior for sparse input neural radiance fields, arXiv preprint arXiv:2305.00041.

[19] J. Wynn, D. Turmukhambetov, Diffusionerf: Regularizing neural radiance fields with denoising diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 4180–4189.

[20] W. Bian, Z. Wang, K. Li, J.-W. Bian, V. A. Prisacariu, Nope-nerf: Optimising neural radiance field with no pose prior, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 4160–4169.

[21] C. Yang, P. Li, Z. Zhou, S. Yuan, B. Liu, X. Yang, W. Qiu, W. Shen, Nerfvs: Neural radiance fields for free view synthesis via geometry scaffolds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 16549–16558.

[22] D. Lee, M. Lee, C. Shin, S. Lee, Dp-nerf: Deblurred neural radiance field with physical scene priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 12386–12396.

[23] T. Hu, X. Xu, S. Liu, J. Jia, Point2pix: Photo-realistic point cloud rendering via neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 8349–8358.

[24] Z.-X. Yin, J. Qiu, M.-M. Cheng, B. Ren, Multi-space neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 12407–12416.

[25] T. Wu, J.-M. Sun, Y.-K. Lai, L. Gao, De-nerf: Decoupled neural radiance fields for view-consistent appearance editing and high-frequency environmental relighting, in: ACM SIGGRAPH 2023 conference proceedings, 2023, pp. 1–11.

[26] X. Zhang, A. Kundu, T. Funkhouser, L. Guibas, H. Su, K. Genova, Nerflets: Local radiance fields for efficient structure-aware 3d scene representation from 2d supervision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 8274–8284.

[27] F. Rivas-Manzaneque, J. Sierra-Acosta, A. Penate-Sanchez, F. Moreno-Noguer, A. Ribeiro, Nerflight: Fast and light neural radiance fields using a shared feature grid, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2023, pp. 12417–12427.

[28] C. Zeng, G. Chen, Y. Dong, P. Peers, H. Wu, X. Tong, Relighting neural radiance fields with shadow and highlight hints, in: ACM SIGGRAPH 2023 Conference Proceedings, 2023, pp. 1–11.

[29] G. Metzer, E. Richardson, O. Patashnik, R. Giryes, D. Cohen-Or, Latent-nerf for shape-guided generation of 3d shapes and textures, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 12663–12673.

[30] C. Deng, C. Jiang, C. R. Qi, X. Yan, Y. Zhou, L. Guibas, D. Anguelov, et al., 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, 2023, pp. 20637–20647.

[31] Y. Jiang, P. Hedman, B. Mildenhall, D. Xu, J. T. Barron, Z. Wang, T. Xue, Alignerf: High-fidelity neural radiance fields via alignment-aware training, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 46–55.

[32] X. Huang, W. Li, J. Hu, H. Chen, Y. Wang, Refsr-nerf: Towards high fidelity and super resolution view synthesis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 8244–8253.