IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 March 2024, accepted 21 March 2024, date of publication 25 March 2024, date of current version 1 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3381531

RESEARCH ARTICLE

Novel View Synthesis and Dataset Augmentation for Hyperspectral Data Using NeRF

RUNCHUAN MA[1], TENGFEI MA[1], DEYU GUO[1], AND SAILING HE[1,2], (Fellow, IEEE)
[1]State Key Laboratory of Modern Optical Instrumentations, Centre for Optical and Electromagnetic Research, Zhejiang University, Hangzhou 310058, China
[2]School of Electrical Engineering, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

Corresponding author: Sailing He (sailing@kth.se)

This work was supported in part by Ningbo Science and Technology Project 2021Z076 and Project 2020Z077, in part by the Key Research and Development Program of Zhejiang Province under Grant 2022C03051 and Grant 2021C03178, in part by the Special Development Fund of Shanghai Zhangjiang Science City, and in part by the National Natural Science Foundation of China under Grant 11621101.

ABSTRACT Hyperspectral data for the 3D domain is relatively difficult to acquire. Existing hyperspectral datasets are unsuitable for 3D research, suffer from issues of severe data scarcity, and a lack of multi-perspective images of the same object, etc. To address these challenges, data augmentation with limited data is essential. In this study, we applied neural rendering method (such as Neural Radiance Field) to hyperspectral images for dataset augmentation. We conducted experiments on novel view synthesis for hyperspectral images from 360-degree multi-perspectives, demonstrating that our method can generate high-quality hyperspectral images from various perspectives. Through experiments involving key points extraction and 3D reconstruction, we validated the efficacy of generating a substantial volume of high-quality hyperspectral images from a restricted set of varying perspectives. These results contribute to addressing the challenges associated with data augmentation. We also conducted experiments of neural radiance fields in the hyperspectral data domain under different network parameters and training conditions to find the appropriate settings.

INDEX TERMS Dataset augmentation, hyperspectral image, NeRF, novel view synthesis.

I. INTRODUCTION

Hyperspectral images contain optical information across multiple spectral bands, far surpassing RGB images in terms of information content. Spectral imaging (see e.g. [2], [3], [4], [5]) combining with AI [6], [7] can give many interesting applications such as remote sensing [8], [9] and detection [10], [11].

In recent years, the 3D field has witnessed the emergence of many excellent technologies, which have been applied to image-based 3D reconstructions or novel view synthesis. 3D reconstruction based on multi-view images, including Structure from Motion (SfM) [12], Neural Radiance Fields (NeRF) [1], and their subsequent developments, brings objects and scenes abundant in two-dimensional images into the realm of three dimensions. The development of image-based 3D technologies has driven progress in various fields like pose estimation [13], [14], [15], 3D modeling [14], [16], [17], and AR&VR [16], and it continues to have enormous growth potential.

The combination of hyperspectral data with 3D offers significant application prospects. Most research in 3D and deep learning [14] currently focuses on data captured by RGB cameras, while hyperspectral images for 3D have not received enough attention. The primary reason is that, compared to RGB cameras, acquiring hyperspectral cameras is challenging, and capturing hyperspectral images is more difficult. For instance, many hyperspectral cameras obtain different spectral bands through filtering light waves of different wavelengths, which require much more times captures than RGB cameras. Additionally, due to the significant reduction in light intensity caused by filtering, longer exposure times are needed to capture than RGB cameras. Consequently, collecting hyperspectral images requires much more time than

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

RGB images. Spectrometer-based hyperspectral cameras can only image one line at a time, which results in slow capture speeds and presents issues related to stability and stitching, all of which affect image quality. Spatial encoding hyperspectral cameras, while capable of fast imaging, have lower spatial and spectral resolutions [18].

While these issues do not directly limit the application of hyperspectral images in the 3D domain, they severely affect the richness of hyperspectral datasets. Currently, we have not found suitable publicly available datasets for multi-view 3D research. This situation indirectly results in a lack of attention to hyperspectral data in the 3D domain. For instance, pose detection requires a large amount of data from different perspectives for training, SfM needs enough images from various perspectives to obtain a dense key point cloud, and neural radiance fields, due to the challenges of capturing a 360-degree hyperspectral dataset, have remained focused on face forward rendering. So far, compared to 3D research using RGB data, 3D research based on hyperspectral data is relatively scarce.

We propose a 3D data augmentation method based on hyperspectral data. With a limited amount of hyperspectral data, we use a neural rendering technology, hyperspectral neural radiance field, adapted from neural radiance field (NeRF) [1] technology for data augmentation, obtaining high-quality hyperspectral data from multiple perspectives to meet the needs of 3D-related research. Data augmentation primarily requires that the generated dataset be of high quality, preserving most of the three-dimensional structural information without distortions or confusions. This is fundamental for conducting 3D research. Additionally, it should retain most of the high-frequency details, such as text and shadows, which could be beneficial for subsequent tasks. Finally, the intensity of light in different spectral bands should be correctly preserved to satisfy the needs for hyperspectral information in subsequent tasks. Currently, there is limited research on NeRF in the hyperspectral domain [19], and it has primarily focused on face forward rendering rather than 360-degree multi-perspective rendering. The number of hyperspectral channels and resolution is also inadequate to meet the requirements of data augmentation scenarios. Through numerous experiments, this paper demonstrates that NeRF-based hyperspectral data augmentation can meet the data augmentation requirements under suitable parameters.

A. MAIN CONTRIBUTIONS

Main contributions of this paper are listed below:

1. This paper experimentally verifies that NeRF can be used for dataset augmentation of hyperspectral data through novel view synthesis.

2. This paper is the first to conduct experiments on 360-degree multi-view synthesis of hyperspectral data using NeRF technology.

3. This paper extends the processing capabilities of NeRF for hyperspectral data to 34 channels and a resolution of 640×480, surpassing previous work in this area.

4. This paper adjusts and experiments with factors such as dataset size and NeRF neural network parameters to obtain the optimal configuration for the current task.

II. RELATED WORK

In this section, the paper will introduce existing relevant research, primarily divided into the following parts: 3D methods like Neural Radiance Fields (NeRF) [1] and Stereo from Motion (SfM), Hyperspectral 3D reconstruction.

A. NERF AND SFM

Neural Radiance Fields primarily work on novel view synthesis. NeRF [1] is currently the most effective method in this field, with their proposal and subsequent research receiving considerable attention. Methods related to Neural Radiance Fields use a certain number of images from different perspectives as input. After training the neural network, they can generate images of objects from any perspective. This method has gained attention because the generated images closely match the original images, allowing for realistic rendering of object details, shadows, lighting, and color information. Neural Radiance Fields do have some issues, such as requiring consistency in input information like lighting and object pose among the different perspective images to ensure correct convergence. Training also takes a considerable amount of time, as does rendering. Neural Radiance Fields have seen numerous derivative works, such as NeRF in real world outdoor scenes [20], which aim to address inconsistencies in information across different time scenarios. There are also efforts to improve NeRF's training speed [21], [22], support for dynamic scenes [23], [24], image resolution [25], model reconstruction [26], etc.

SfM is a more traditional kind of 3D reconstruction method. Despite its long history, it still holds a crucial position in the 3D domain. SfM also generates 3D information of objects from different viewpoints. Common methods in this field are based on key point recognition and matching for 3D reconstruction. Traditional key point recognition methods like LIFT [27], SIFT [28], etc., have been tested over the years. With the increased application of deep learning methods, key point recognition based on deep learning has received much attention in recent years, for example, Superpoint [29], and has made significant contributions to this field. There are also other methods like Time-Domain methods [30] in this area.

B. HYPERSPECTRAL-BASED 3D

Despite the vibrant development in the 3D domain, hyperspectral data have not received as much attention as RGB images. In the domain of 3D reconstruction based on SfM, articles like the one by Ma et al. [31] and others have contributed to key point recognition and matching for hyperspectral images. They trained neural networks on hyperspectral images and obtained far more key points than RGB images, resulting in denser point clouds and higher reconstruction accuracy. However, the issue of discrete key

points has not been completely resolved, as discussed later. This method still faces challenges in correctly handling surfaces lacking texture in new viewpoint images.

Another related field is active 3D hyperspectral imaging. This method requires the use of additional equipment to project onto the target, enabling the hyperspectral camera to acquire three-dimensional information, as discussed in articles such as those by Luo et al. [32] and others. While this method allows end-to-end acquisition of three-dimensional information, it has higher equipment requirements and usage conditions compared to fully image-based methods.

In the field that combines hyperspectral and NeRF, only one paper [19] was found that describes a method for NeRF reconstruction of cross-band images. This paper proposed a method designed for tackling 14 different types of channels, including RGB images, 10-channel hyperspectral data, and single-channel near-infrared image data. Using NXDC method proposed by themselves, they trained NeRF for face forward scene rendering and achieved good results. However, their method is limited to face forward scenes, which means it only captures narrow perspectives of the object from the front, and it cannot be directly extended to images from all 360 degree perspectives, which is essential for data augmentation task. Additionally, their hyperspectral image resolution is only 254×510, with 10 channels. There is no evidence to suggest that their method can achieve the same quality of reconstruction for high-resolution (640×480) full visible spectrum hyperspectral data (with more than 30 channels). A new method is necessary to accomplish data augmentation task.

This paper aims to fill the gaps in the above research, confirming the effectiveness of Neural Radiance Fields in hyperspectral images and further demonstrating that using this method for multi-perspective dataset augmentation of hyperspectral images is effective.

III. METHOD OF NOVEL VIEW SYNTHESIS BASED ON NEURAL RENDERING

The method used in this paper involves extending the neural radiance field into the hyperspectral domain. It trains the neural network with a certain number of hyperspectral images taken from different perspectives to obtain a model capable of rendering images from any viewpoint. By rendering a significantly greater number of high-quality hyperspectral images than the original data, it augments the hyperspectral dataset. This section will be divided into two modules, including details of the NeRF technology we employed and information about the hyperspectral images used in the experiments.

A. NOVEL VIEW SYNTHESIS BASED ON NEURAL RADIANCE FIELDS

NeRF [1] is a neural network-based novel view synthesis technology. Its purpose is to train a neural network to implicitly represent the three-dimensional distribution density and color information of the target object. Based on this three-dimensional distribution information, it samples rays from corresponding perspectives, rendering images from any viewpoint using a camera model.

Neural Rendering: Specifically, we first define a neural network structure, which consists of 8 fully connected layers and corresponding output layers for distribution density and intensity in different channels. The input to the neural network is a series of coordinates, which are processed through the neural network to obtain spatial density distribution and color information at each coordinate. When generating an image from a particular viewpoint, pixel values are rendered pixel by pixel. The value of each pixel is determined by the density distribution and color information at a series of coordinates sampled along the ray defined by the camera model, using the following formula:

$$\overline{C}(r) = \sum_{i=1}^{N} T_i(1 - exp(-\sigma_i\delta_i))c_i, T_i = exp(-\sum_{j=1}^{i-1}\sigma_j\delta_j) \quad (1)$$

Here, $\overline{C}(r)$ represents the color information of the pixel corresponding to the ray r, which is a three-dimensional vector in RGB images, or has a dimension equal to the number of hyperspectral channels in hyperspectral images. c_i represents the color information of the material at the i-th sampling point, with a dimension matching the number of hyperspectral channels. T_i indicates the probability that light from the camera center can reach this point without obstruction at the i-th sampling point. σ_i represents the spatial density distribution function at the i-th sampling point, and δ_i represents the distance interval at the i-th sampling point. The formula shows that each pixel is rendered based on the spatial density distribution and color information at the corresponding sampling points along its ray path. To generate a complete image, every pixel needs to be rendered sequentially.

Position Encoding: NeRF [1] doesn't directly input three-dimensional coordinates into the neural network when dealing with input coordinates. Instead, it uses position encoding:

$$\gamma(p) = (sin(2^0\pi p), cos(2^0\pi p), .., sin(2^{L-1}\pi p), cos(2^{L-1}\pi p)) \quad (2)$$

In this equation, p is the coordinate value of a given point, L is the frequency level. This is because neural networks tend to prioritize fitting low-frequency data and may not fit high-frequency data well. Position encoding improves the network's ability to represent high-frequency data effectively, thereby enhancing the training of the neural network. In our experiment, the value of L for the position encoding used during training is 10.

Training: During neural network training, pixels of training sample images are sampled, and a series of coordinate values on the rays generated by the camera model are input into the neural network. This yields density distribution and color information at the corresponding coordinates. As a result, the neural network's output value for that pixel is obtained. Loss functions are computed based on the difference between this

value and the ground truth value, and error backpropagation is used for training. In this paper uses hyperspectral data to train the neural network. The color information output by the neural network is adjusted to 34 channels to render values for 34 different channel images.

Network Structure: The main body of NeRF's neural network consists of fully connected layers. In the following experiments, the parameter adjusted in the network structure is the number of neurons in the fully connected layers, which is referred to as its width. Extensive experiments were conducted to evaluate the image quality of the neural network output for various widths.

B. RELEVANT DETAILS OF HYPERSPECTRAL IMAGES

The hyperspectral images used in this paper have 34 channels with wavelengths distributed from 420nm to 750nm at 10nm intervals. The image resolution is 640×480. The images were captured in advance [31]. The experiment required capturing hyperspectral images from various perspectives of the same target. An electric turntable was chosen for stable image capture. Although using turntable does not completely adhere to NeRF's acquire of a stationary object and lighting environment with only the camera revolving around it, it does not significantly impact the main part of the target, as we shall see. Like many hyperspectral image capturing methods, the approach used in this experiment requires capturing images for each channel separately. Due to the low light transmission after filtration, each channel requires an extended exposure time to capture clear images. This can result in several minutes of imaging time for a complete hyperspectral image from a single perspective. Even so, the results in the near-infrared and ultraviolet parts of the spectrum may still appear darker. Hence, using a turntable to automate photography rather than holding cameras by hands was the only viable choice.

We chose the scene used in the experiments from various experimental scenarios, and this scene has a wide range of features, allowing us to evaluate the effect of NeRF on hyperspectral data augmentation under various conditions. First, the images contain rich high-frequency information, which is well-represented in various frequency bands. This can demonstrate whether this method can preserve image details. Second, the images include smooth areas with textures missing, which are challenging for traditional SfM methods to model. A specific comparison will be made in the results section. The image comparisons can also demonstrate whether our method yields good results in dealing with regions lacking texture. Finally, the upper part of the images contains inconsistent information due to the change of lighting conditions, such as shadows that do not move with the object. The experiment will verify whether a small amount of inconsistent content can lead to the loss of a large amount of image information and will assess the method's robustness using hyperspectral images, further demonstrating its resistance to spatial interference, which is essential when capturing long-duration hyperspectral images.

IV. EXPERIMENT AND RESULT

A. EXPERIMENT DESIGN OF HYPERSPECTRAL DATASET AUGMENTATION BASED ON NOVEL VIEW SYNTHESIS

In the results section, we conducted a total of four series of experiments to demonstrate the effectiveness of our method in dataset augmentation and to explore the outcomes of training with different parameters. Ultimately, we determined the parameter settings suitable for dataset augmentation tasks. In this section, we first present the experimental results of the hyperspectral NeRF (Neural Radiance Fields). We then validate the effectiveness of our method for dataset augmentation through SfM (Structure from Motion) three-dimensional reconstruction experiments. Finally, we investigate its performance under different dataset sizes and network widths.

The data used to train the model in this experiment was collected using a hyperspectral camera, which captures images in various wavelength bands ranging from 420 to 750nm, with a 10nm interval, resulting in a total of 34 spectral channels. The images captured by the 8-bit camera have a resolution of 640×480 pixels, and in this experiment, we used the original resolution for training. A 100w LED light source was used during the collection of hyperspectral images. The dataset we use includes a total of 48 perspective images, with each perspective distributed around the target object in a full circle, and the minimum angle between any two perspectives was 6 degrees, with some perspectives having a 12-degree separation. In total, 48 different perspectives of hyperspectral images were previously captured, and they were subsequently divided into training and test sets for later training. The training parameters in this experiment were set with a batch size of 32,768 for ray sampling, with a network width varies from 8 to 512, and with a training set size of 12, 24, 36, 40, 42 to test the performance. When the network is 512 wide, our GPU ran out of memory, so we have to half the batch size to 16,384, which lowers the performance a little bit. Other parameters were consistent with NeRF [1]. The model was trained for at most 200,000 epochs on an NVIDIA 2080TI 11G GPU, with each training session taking approximately 2-17 hours, depending on different width of networks.

B. EXPERIMENTAL RESULTS OF HYPERSPECTRAL NERF

In the results section, this paper will show the effects of dataset augmentation with the neural radiance field under different conditions. First, we will assess the model's generation performance in training with 34-channel hyperspectral images at 42 different perspectives to show the performance of our method. Next, we will examine the impact of varying training set sizes on the results, followed by an analysis of the influence of different network width. After we find the best parameters for dataset augmentation, an experiment on SfM reconstruction will finally be conducted to demonstrate the applicability of NeRF in dataset augmentation within the hyperspectral domain.

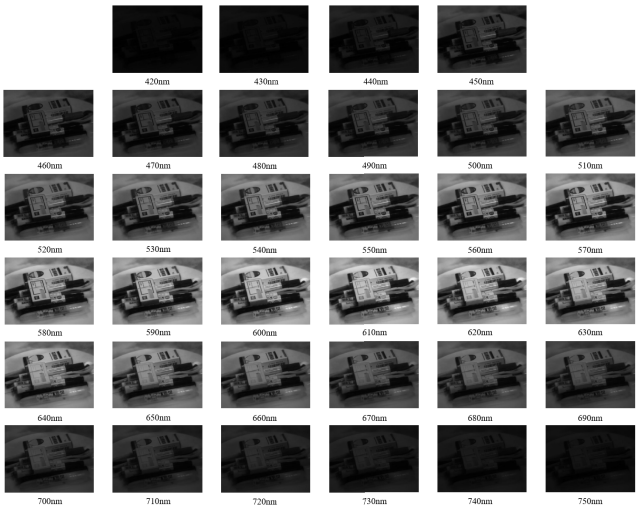


FIGURE 1. Some of our training results. For each perspective, images of 34 channels are generated, corresponding to 34 spectrum ranges (the central wavelengths are 420nm, 430nm, ..., 750nm) with a band width of 10nm.

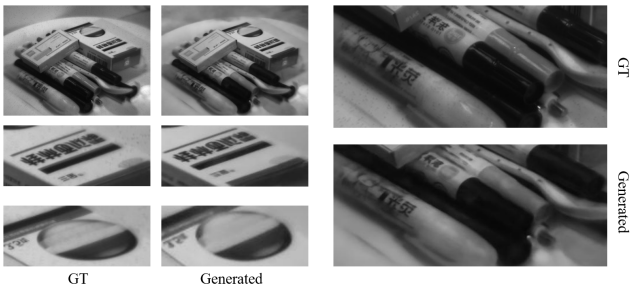


FIGURE 2. Comparison between the generated images (Generated) and the ground truth (GT) at 600nm wavelength. Full images (left top) and zoom-in images (left bottom and right) are illustrated. In most regions, there is no significant difference between the generated images and the ground truth.

1) DATA AUGMENTATION OF HYPERSPECTRAL IMAGES WITH NERF

This study initially investigates the applicability of NeRF in the field of hyperspectral images. In this experiment, to get the best performance of hyperspectral NeRF, we trained a neural network by inputting hyperspectral images from 42 different perspectives. In this experiment, we selected 6 different perspectives as a test set. To assess data augmentation, the neural network rendered a total of 120 images from different perspectives, each separated by 3 degrees. Some of the training results are shown in Fig.1, while details are shown in Fig.2.

From the perspective of image details, the quality of the generated images is quite high, preserving most of the fine details, including text, shadows, and channel intensities. Although smaller text appears slightly blurry, and some reflections on the right side were not fully restored, overall, there are no significant differences between the generated images and the original images.

The quality of images needs to be quantified using standardized metrics, and the commonly used metric is the

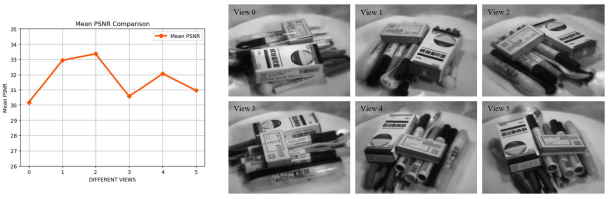


FIGURE 3. PSNR value of generated images from different perspectives (LEFT), and the corresponding generated images at 600nm wavelength (RIGHT). SSIM values of the corresponding images are given in the appendices.

TABLE 1. Comparison of results from different methods [1].

Method	PSNR	Rank
SRN-RGB	22.26	-
NV-RGB	26.05	-
LLFF-RGB	24.88	-
NeRF-RGB	31.01	Second
Hyperspectral-NeRF(ours)	31.67	Best

Peak Signal-to-Noise Ratio (PSNR). We measure the quality of the results by calculating the PSNR value between the rendered result images and the corresponding perspective images in the test set.

$$PSNR = 10 \times log_{10} \frac{MAX_I^2}{MSE} \tag{3}$$

MAX_I in this equation means the maximum value of the current image format, for example 256 in 8-bit images. In this paper, image data are normalized with a maximum value of 1, so $PSNR = -10log_{10}(MSE)$. Larger PSNR value means better quality for images in most cases. The experimental results are shown in the Fig.3.

In this section, six different perspectives generated during the experiment are selected for presentation. The average PSNR value obtained by comparing images generated by the neural network with real images (GT) is 31.67. The PSNR values from different perspectives exhibit a minor fluctuation within ± 1 around the average PSNR of the 6 generated images, and all of them have PSNR values above 30, indicating their high quality.

The image results for each perspective are as shown in Fig.3. The generated images at all perspectives exhibit consistent reproduction of both the subject and fine details. The fluctuations in PSNR values among different perspective images are primarily due to inconsistencies in the top background of the images, leading to incorrect convergence, and improper handling of reflections in certain areas by the neural network.

Given the absence of examples for full 360-degree reconstruction of hyperspectral images in previous research, we compare our results with the results of synthetic RGB images in NeRF.

Table 1 contains PSNR values for the results mentioned above. When compared to NeRF's 360-degree RGB synthesis results, the PSNR value for hyperspectral synthesis is the highest, but it does not significantly surpass the NeRF method

under RGB conditions. However, the hyperspectral NeRF synthesis results far outperform those of non-NeRF methods.

Since no previous results can be found in 360-degree dataset augmentation for hyperspectral images using NeRF, we compare our results with previous RGB view synthesis results as a comprehensive practice. Although RGB images have only 3 channels while hyperspectral images in the present paper have 34 channels, a comparative PSNR of our images shows that our method generated high quality hyperspectral images as NeRF did for RGB images. For both RGB images and hyperspectral images, PSNR values are first computed for each channel and then the mean values across channels are calculated. The meaning of this metric does not change with the total number of channels.

For dataset augmentation, the quality of reconstructed images is of paramount importance, requiring the preservation of pattern details, shadows, and other distinct features. In this regard, methods based on NeRF exhibit a significant advantage over other approaches.

Currently, the primary methods for novel view synthesis, apart from NeRF, are based on 3D reconstruction, such as SfM (stereo from motion) methods. In previous paper [31] of T.Ma et al., 3D reconstruction work based on deep learning key point matching outperformed traditional non-deep learning key point extraction results. We compare the results of this experiment with those from Ma et al. [31], as shown in Fig.4.

In Fig.4, (a) and (b) represent results generated by the SfM method, after key point extraction and dense reconstruction. (c) shows the results rendered by the method used in this experiment, and (d) is the original image. Several advantages of our research method compared to the images generated after SfM reconstruction are evident.

Firstly, results generated by SfM methods based on key points exhibit a noticeable grainy texture. This is due to the sparse nature of their key points. Even if efforts are made to densify the results through interpolation and other methods, since the key points are too sparse, a significant amount of details of the original objects are lost, as we can see in Fig.4 (a) and (b). In contrast, the method based on neural radiance fields has a distinct advantage in this regard. A comparison between (a1) and (c1) shows that the results reconstructed in our paper (c1) significantly enhance details, making it possible to clearly identify text and patterns below, which are lost in (a1).

Secondly, our method can accurately model smooth, textureless regions, whereas SfM methods based on key point matching struggle due to the difficulty of capturing sufficient key points in such areas and issues related to incorrect matches, leading to sparsity and substantial noise. In (a1), there is a hollow section in the upper text and the middle of the pattern below (depicted in blue). When rendering new perspective images, it's challenging to find appropriate pixel fillings, resulting in a background color display. This is caused by the failure to capture key points. In contrast, the images rendered by our model are consistent with the original image, displaying a smooth white surface, a feature that is maintained across all 34 hyperspectral channels.

Furthermore, our experimental model demonstrates a high degree of robustness in the presence of inconsistencies in the background. During the training of the neural network, if different perspectives' images exhibit inconsistencies, it can lead to the neural network's inability to find the object's three-dimensional spatial distribution that satisfies all perspective images, resulting in the convergence to incorrect results. This experiment shows that even when inconsistencies exist in the background, the image of the main target remains unaffected.

The results above demonstrate that our method not only accurately preserves the correct 3D view information for the object across all channels but also precisely retains details such as shadows and text, making it suitable for use as an augmented dataset.

2) IMPACT OF TRAINING SET IMAGE QUANTITY ON RECONSTRUCTION RESULTS

In this section, we conducted an experimental study on the influence of different training set sizes on hyperspectral reconstruction results. In this experiment, the width of the neural network was uniformly set to 256. The total size of the dataset collected for this experiment is 48, and the sizes of the training sets were set to 12, 24, 36, 40 and 42 images, respectively. The perspectives used in the training sets were evenly distributed, whereas images from other perspectives were used as the test set to assess the quality of the training results. The reconstruction results of models trained with different training set sizes are shown in Fig.5.

Results in Fig.5 indicate that when the training set size is greater than 24 different perspectives, there is no significant difference in the training results, and their PSNR values all exceed 30. However, when the training set is reduced to 12 different perspectives, there is a noticeable decrease in the quality of the training results. The comparison of different generated images is shown in the following Fig.6.

In Fig.6, results of rendering images from similar perspectives to the training set are presented for different training set sizes. When the model is trained with training sets containing 42, 40, 36 different perspectives, the image outputs closely match the real images, with only slight blurriness in the upper background due to inconsistencies in the environment during capture, as seen in the comparisons for 42-a, 40-a, and the ground truth (GT). When the training set size is reduced to 24, most parts in the image remains consistent with ground truth, and the text is clearly readable. But when the training set size is reduced to 12, the entire image exhibits distortion and blurriness, making it impossible to recognize the text. Performance like this is obviously not qualified to be used for data augmentation.

These observations align with the description based on PSNR values in Fig.5. When the training set size is greater than or equal to 24, the PSNR values remain above 30, with sufficient clarity to read text. However, when the training set

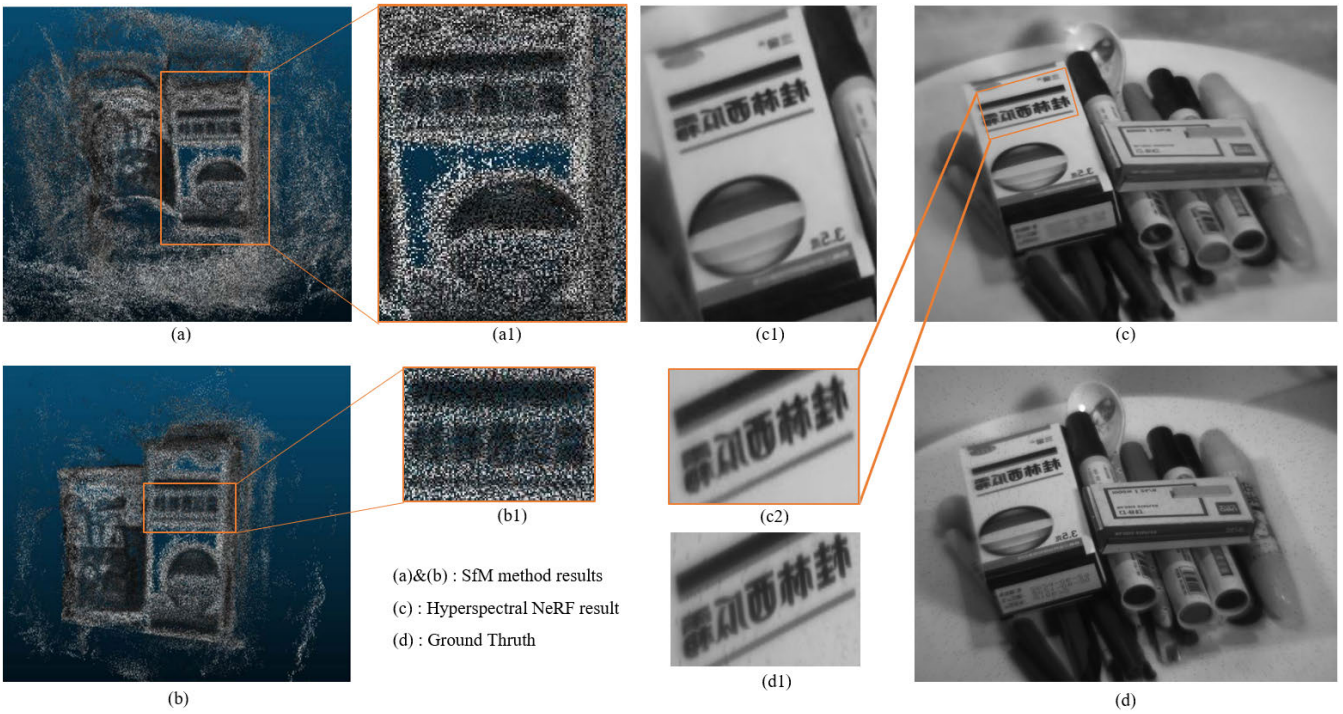


FIGURE 4. (a)-(c) Comparison of the results from different methods, (d) ground truth (GT) images at 600nm wavelength. SfM method results have lower quality then results from our method.

Dataset Size	PSNR	Difference (%)
42	31.7	-0.3%
40	32.5	+2.2%
36	32.1	+0.9%
24	31.8	0%
12	20.4	-35%

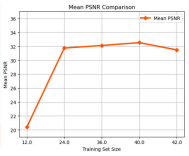


FIGURE 5. Table: Impact of different training set sizes on the performances PSNR and difference (LEFT); Mean PSNR as the training set size increase (RIGHT). SSIM of the corresponding training set sizes are in appendices.

size is reduced to 12, the PSNR value drops to around 20, indicating a significant overall decline in image quality, resulting in widespread blurriness and a foggy distribution in the images. The experimental results also reveal that if the training set contains two perspectives with significant angular differences, the model's performance in generating images between these two widely spaced perspectives is worse compared to generating images between perspectives with smaller differences in the training set.

These experimental results indicate that in the case of using hyperspectral images, NeRF is sensitive to the dataset size and distribution. When the training set is relatively small and the perspectives between the two training perspectives are large, the neural network's representation of the 3D distribution significantly differs from the true values, resulting in bad performance. Based on the experiments, it is evident that in order to generate images qualified for data augmentation, the minimum number of captured perspectives should be at least around 24, and these perspectives should be evenly distributed between 0 and 360 degrees.

3) IMPACT OF NEURAL NETWORK PARAMETER QUANTITY ON RECONSTRUCTION RESULTS

The objective of this experimental section is to investigate the effect of the number of parameters on the reconstruction of hyperspectral images by adjusting the network width used in the NeRF neural network. The basic structure of the NeRF neural network is a fully connected network, and the number of parameters in this network is primarily determined by the network width (the number of neurons in the hidden layers). This experiment selected neural networks with different widths, such as 512, 256, 128, 64, and 32, and conducted experiments using a training set of 24 images. The experimental results are listed in Fig.7.

The experimental results in Fig.7 indicate that PSNR shows an increasing trend with the increase in neural network width, as expected. Even though hyperspectral images contain significantly more information than RGB images, NeRF does not require larger number of parameters to process them. Even when the neural network width is as low as 32, the PSNR value is around 30, which suggests that the main structure of the image is correct. However, a substantial amount of detailed information is lost, which can also be observed from the comparison with 32-1 in Fig.7. As the number of neural network parameters increases, details in the images continuously improve, but as the neural network width approaches 256, the effect enhancement tends to saturate.

The experiment demonstrates that network width has a significant impact on the PSNR of generated images. Considering that the influence of neural network width on

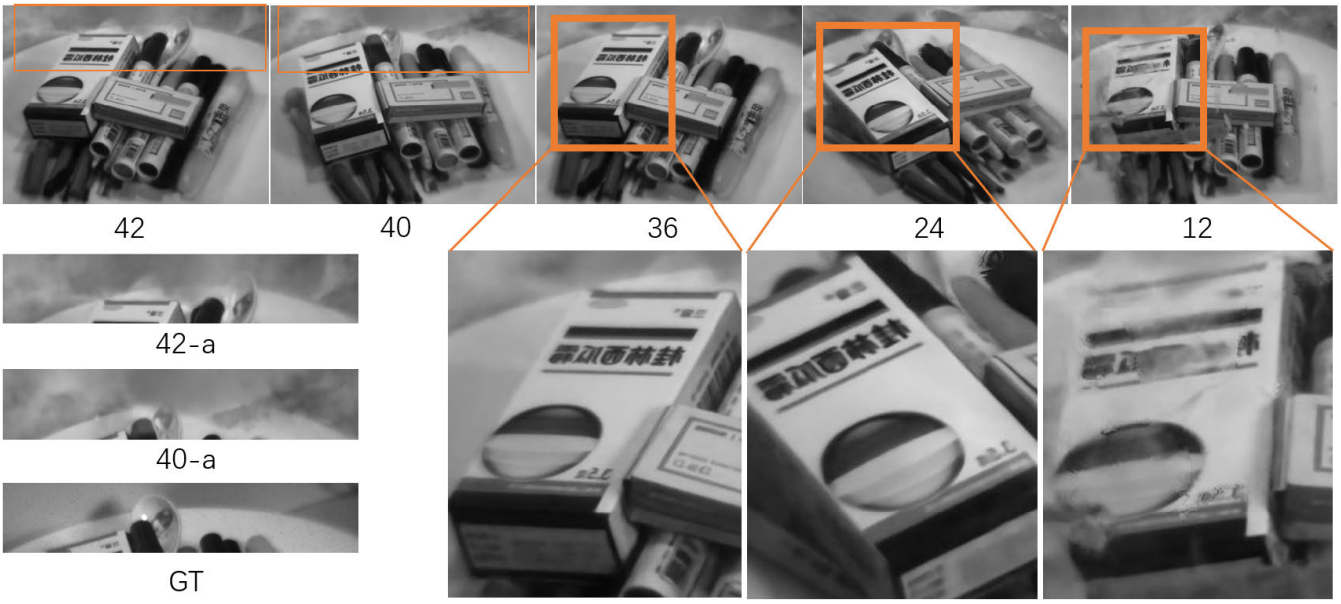


FIGURE 6. Comparison of images generated by different training set sizes at 600nm wavelength. The left bottom shows the ground truth (GT). The training set size increases from 12 to 42 different perspectives. Numbers under the top row of the figure indicate the training set size (42 means 42 perspectives of images was used to train the network). Images on the bottom are zoom in of the images on the top.

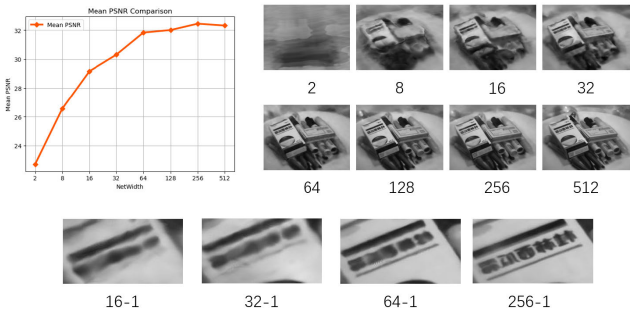


FIGURE 7. Comparison between the images generated by different network widths (from 2 to 512 neurons per layer) at 600nm wavelength. The PSNR plot for each network width is shown (UPLEFT). The example images of each network width from 2 to 512 are shown (UPRIGHT). Some details are shown by zoom in images of the example images (from 16-1 to 256-1) in the bottom row.

TABLE 2. Growth of parameter number vs PSNR growth.

Network Width	Parameter Growth	PSNR	PSNR Growth
32(As Base)	0%	30.3	0%
64	400%	31.8	+5.0%
128	1600%	32.0	+5.6%
256	6400%	32.4	+7.0%
512	25600%	32.3	+6.6%

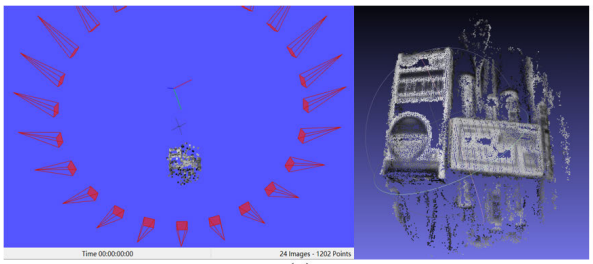
the number of parameters in the fully connected layer is quadratic, an increase in network width results in a significant decrease in training speed. When the width exceeds 256, increasing the neural network width has a limited effect on the performance improvement, and their relationship is given in Table 2. The data corresponding to a network width of 32 are selected as the baseline, and the other data are compared to it.

From the information in Table 2, it is evident that as the neural network width increases, the improvement in NeRF's novel view synthesis tends to saturate. Furthermore, in cases of excessive width, it is prone to lead to overfitting of the neural network. In our case this phenomenon is significant only when the width is set to 512, still indicating that further increasing the neural network width does not bring any benefit.

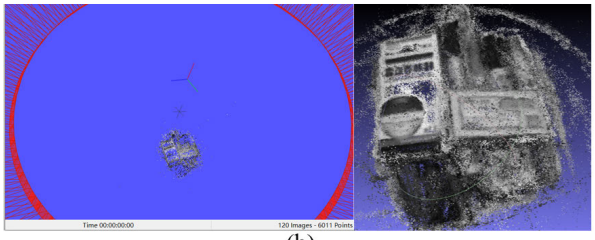
The experimental results demonstrate that, even though hyperspectral image channels are several times greater than ordinary RGB images, due to the expressive capabilities of fully connected layers, NeRF neural networks can maintain good training results within a significant parameter range. Among them, the width of 256 provides the optimal performance, and it is recommended to choose this width for dataset augmentation. When the number of hyperspectral image channels changes, this optimal parameter may also change accordingly.

4) EFFECT OF DATASET AUGMENTATION ON SFM 3D RECONSTRUCTION

For images that are suitable for 3D related tasks, consistent spacial features are required. 3D reconstruction datasets should contain highly consistent multi perspective images of the same object. It is hard to achieve this for hyperspectral images because it's much slower to obtain hyperspectral imaging as compared with RGB images. To demonstrate the effectiveness of dataset augmentation in 3D related tasks, consistent features of the object in the generated images shall be shown. Thus an experiment is conducted with the best fitting parameters introduced above, to show that images generated by our method can preserve enough local features for key point recognition and 3D reconstruction.



(a)



(b)

Data Augmentation	Points Found	Dense Reconstruction
No	~1200	Sparse
Yes	~6000	Dense

(c)

FIGURE 8. Comparison of 3D reconstruction results before and after dataset augmentation. (a) Result before augmentation. (b) Results after Augmentation. (c) Details and comparison.

In this experiment, we choose 24 perspectives images as original data set. Hyperspectral NeRF with network width 256 is trained with the original data set for augmentation, which generates 120 images with different perspective. This experiment conducts 3D reconstruction using both the original data (24 perspective images) and the augmented data (120 perspective images) with traditional key point recognition method. Sparse and dense reconstructions are carried out using COLMAP [33], [34], and the results are shown in Fig.8:

In Fig.8a, the result of reconstruction without dataset augmentation is shown. In the original 24 images, a total of 1202 key points are found. Even after subsequent dense reconstruction, it is evident that the key points extracted are relatively sparse. Not only are the smooth surfaces not adequately modeled, but also only a small fraction of the feature-rich objects is captured.

Fig.8b displays the results of reconstruction using 120 images after dataset augmentation. In this experiment, we find over 6000 key points in total. After dense reconstruction, the key point density is higher, and the primary features of all objects have been correctly captured. Only some points on the smooth surface are missing. The results have significantly improved compared with the previous version.

Local spatial features are extremely important in key point collection and recognition tasks, which are typical kinds of 3D works, broadly used in areas like computer vision, slam, 3D reconstruction, etc. Number of key points collected can illustrate whether the local features of the object are preserved well enough for 3D related tasks.

The experiment demonstrates that a large number of high-quality hyperspectral images generated by the hyperspectral NeRF method in this paper reserves adequate amount of local features, and can bring substantial improvements to works in the 3D domain when used for dataset augmentation.

C. DISCUSSION

The purpose of this paper is to demonstrate the suitability of NeRF for 360-degree multi-view reconstruction in the hyperspectral domain and its' remarkable performance in data augmentation. The result above showed that the novel view synthesis technology based on NeRF can expand hyperspectral image datasets for reconstruction and other 3D applications, thereby enriching research and applications of hyperspectral data in the 3D domain. The experiments successfully generated 120 perspective images with high quality from at least 24 different perspective hyperspectral images. In the generated image results, the primary objects in the images have not undergone distortions that affect their structure. Even when there are inconsistencies in the shadows in the background and noise in the images, under appropriate model parameter settings and dataset sizes, the foggy spatial distribution caused by the inability to correctly converge due to background inconsistencies is limited to the vicinity of the inconsistent scene. This does not affect the imaging quality of the primary imaging targets, with PSNR values exceeding 30, signifying high image quality.

The quality of image generation varies among different channels. For multi-channel images, the darker channels, including the violet channel near 400nm and the red channel near 700nm, have higher PSNR values, generally around 40. These wavelength bands have darker images with much less information compared to other bands, making NeRF's reconstruction of them easier. The brighter channels around 500nm have PSNR values ranging from 27 to 35, which, although lower than the darker channels in terms of PSNR, carry a significantly higher amount of data and still maintain relatively high quality.

By adjusting the dataset size, we found that for 360-degree multi-view reconstruction, roughly evenly captured 24 images can generate accurate and highly detailed images at any perspective. This further reduces the requirement for many hyperspectral capture images. Through the experiment in section IV-B4, we demonstrate that our method on dataset augmentation have significant effect on 3D areas like SfM reconstruction, which fulfill our goal in this paper.

In summary, this paper successfully validates the potential of NeRF-based new view synthesis technology for expanding hyperspectral datasets. The results indicate that with acceptable training set sizes and neural network parameter numbers, high-quality imaging is achieved. This can be applied to other 3D-related tasks, especially those requiring 360-degree multi-view datasets.

However, our method still have some limitations. Our method does not work well under datasets with few perspectives (for example, less then 20 perspectives for 360-degree datasets). Our method can not handle moving objects or scenes.

Currently, this research offers several directions for future exploration, such as addressing the decline in imaging quality when dealing with small datasets. Additionally, this research primarily focuses on dataset augmentation for different perspectives and does not consider scenarios with varying lighting conditions. These are areas that warrant further investigation.

V. CONCLUSION

In this study, we conducted multiple experiments to demonstrate the applicability of neural rendering method for novel view synthesis in the hyperspectral domain. We have verified the feasibility of using hyperspectral NeRF for multi-perspective dataset augmentation of hyperspectral data. Our method can successfully generate hyperspectral images at various perspectives, preserve the object's feature information which enabling it to play a role in key point recognition and 3D reconstruction. Our method may still be improved in terms of reduce required training set image quantity, increase generated image quality and other aspects through future research. It is our hope that by augmenting hyperspectral images from various perspectives, hyperspectral imagery can garner greater attention in the field of 3D research and applications.

APPENDIX SSIM METRICS

See Figures 9 and 10.

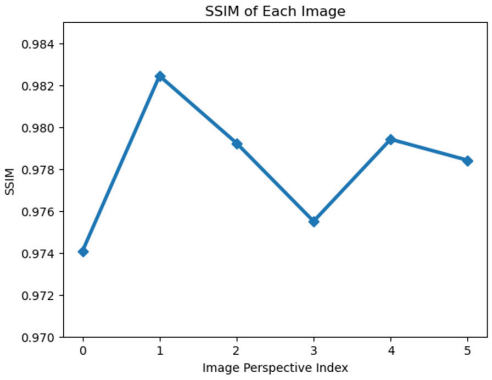
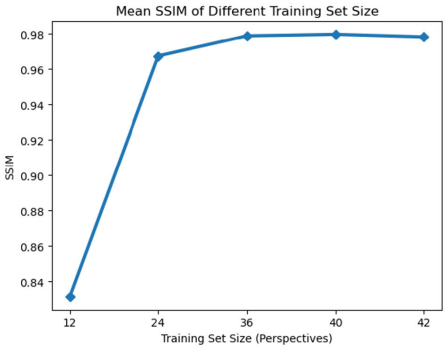


Image Perspective Index	SSIM	Rank
0	0.974	
1	0.982	Best
2	0.979	
3	0.975	
4	0.979	
5	0.978	

FIGURE 9. SSIM data related with the PSNR data in Fig.3.



Training Set Size	SSIM	Rank
42	0.978	
40	0.979	Best
36	0.978	
24	0.967	
12	0.831	

FIGURE 10. SSIM data related with the PSNR data in Fig.5.

REFERENCES

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, Jan. 2022.

[2] C. Zhang, A. Yang, and S. He, "Lateral flow immunoassay strip based on confocal Raman imaging for ultrasensitive and rapid detection of COVID-19 and bacterial biomarkers," *Prog. Electromagn. Res. M*, vol. 120, pp. 41–54, 2023.

[3] H. Zhu, J. Luo, J. Liao, and S. He, "High-accuracy rapid identification and classification of mixed bacteria using hyperspectral transmission microscopic imaging and machine learning," *Prog. Electromagn. Res.*, vol. 178, pp. 49–62, 2023.

[4] C. Jiao, Z. Lin, Y. Xu, and S. He, "Noninvasive Raman imaging for monitoring mitochondrial redox state in septic rats," *Prog. Electromagn. Res.*, vol. 175, pp. 149–157, 2022.

[5] T. Guo, Z. Lin, X. Xu, Z. Zhang, X. Chen, N. He, G. Wang, Y. Jin, J. Evans, and S. He, "Broad-tuning, dichroic metagrating Fabry–Perot filter based on liquid crystal for spectral imaging," *Prog. Electromagn. Res.*, vol. 177, pp. 43–51, 2023.

[6] D. Gong, T. Ma, J. Evans, and S. He, "Deep neural networks for image super-resolution in optical microscopy by using modified hybrid task cascade U-Net," *Prog. Electromagn. Res.*, vol. 171, pp. 185–199, 2021.

[7] H. Wang, D. Gong, G. Cheng, J. Jiang, D. Wu, X. Zhu, S. Wu, G. Ye, L. Guo, and S. He, "Detecting temperature anomaly at the key parts of power transmission and transformation equipment using infrared imaging based on SegFormer," *Prog. Electromagn. Res. M*, vol. 119, pp. 117–128, 2023.

[8] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, p. 2659, Aug. 2020.

[9] P. Shippert, "Why use hyperspectral imagery?" *Photogramm. Eng. Remote Sens.*, vol. 70, no. 4, pp. 377–396, 2004.

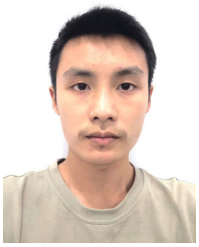
[10] M. Mehdorn, H. Köhler, S. M. Rabe, S. Niebisch, O. Lyros, C. Chalopin, I. Gockel, and B. Jansen-Winkeln, "Hyperspectral imaging (HSI) in acute mesenteric ischemia to detect intestinal perfusion deficits," *J. Surgical Res.*, vol. 254, pp. 7–15, Oct. 2020.

[11] J. Behmann, J. Steinrücken, and L. Plümer, "Detection of early plant stress responses in hyperspectral images," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 98–111, Jul. 2014.

[12] Y. Furukawa and C. Hernández, "Multi-view stereo: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 9, nos. 1–2, pp. 1–148, 2015.

[13] C. Sahin, G. Garcia-Hernando, J. Sock, and T.-K. Kim, "A review on object pose recovery: From 3D bounding box detectors to full 6D pose estimators," *Image Vis. Comput.*, vol. 96, Apr. 2020, Art. no. 103898.

[14] A. Grabner, P. M. Roth, and V. Lepetit, "3D pose estimation and 3D model retrieval for objects in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3022–3031.
[15] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, "INeRF: Inverting neural radiance fields for pose estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 1323–1330.
[16] M.-D. Yang, C.-F. Chao, K.-S. Huang, L.-Y. Lu, and Y.-P. Chen, "Image-based 3D scene reconstruction and exploration in augmented reality," *Autom. Construction*, vol. 33, pp. 48–60, Aug. 2013.
[17] A. M. Intwala and A. Magikar, "A review on process of 3D model reconstruction," in *Proc. Int. Conf. Electr., Electron., Optim. Techn. (ICEEOT)*, Mar. 2016, pp. 2851–2855.
[18] M. Halicek, H. Fabelo, S. Ortega, G. M. Callico, and B. Fei, "In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: Revealing the invisible features of cancer," *Cancers*, vol. 11, no. 6, p. 756, May 2019.
[19] M. Poggi, P. Z. Ramirez, F. Tosi, S. Salti, S. Mattoccia, and L. D. Stefano, "Cross-spectral neural radiance fields," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2022, pp. 606–616.
[20] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "NeRF in the wild: Neural radiance fields for unconstrained photo collections," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7206–7215.
[21] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, "Depth-supervised NeRF: Fewer views and faster training for free," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12872–12881.
[22] C. Sun, M. Sun, and H.-T. Chen, "Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5449–5459.
[23] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-NeRF: Neural radiance fields for dynamic scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10313–10322.
[24] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable neural radiance fields," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5845–5854.
[25] Z. Wang, L. Li, Z. Shen, L. Shen, and L. Bo, "4K-NeRF: High fidelity neural radiance fields at ultra high resolutions," 2022, *arXiv:2212.04701*.
[26] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," 2021, *arXiv:2106.10689*.
[27] K. M. Yi et al., "LIFT: Learned invariant feature transform," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VI 14*. Springer, 2016.
[28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[29] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 224–236.
[30] J. Zhong, X. Zhang, and S. Liu, "Three-dimensional imaging of target based on time-domain sparse representation of multi-view SAR data," *Prog. Electromagn. Res. C*, vol. 138, pp. 145–159, 2023.
[31] T. Ma, Y. Xing, D. Gong, Z. Lin, Y. Li, J. Jiang, and S. He, "A deep learning-based hyperspectral keypoint representation method and its application for 3D reconstruction," *IEEE Access*, vol. 10, pp. 85266–85277, 2022.
[32] J. Luo, Z. Lin, Y. Xing, E. Forsberg, C. Wu, X. Zhu, T. Guo, G. Wang, B. Bian, D. Wu, and S. He, "Portable 4D snapshot hyperspectral imager for fastspectral and surface morphology measurements," *Prog. Electromagn. Res.*, vol. 173, pp. 25–36, 2022.
[33] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4104–4113.
[34] J. L. Schönberger et al., "Pixelwise view selection for unstructured multi-view stereo," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14*. Springer, 2016.



RUNCHUAN MA received the B.S. degree in optics engineering from BIT, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree in optical engineering with Zhejiang University, Hangzhou, China. His research interests include artificial intelligence, NeRF, and 3D image processing.

TENGFEI MA was born in Yantai, Shandong, China, in 1995. He received the B.Eng. degree in communication engineering from Jilin University, Changchun, China, in 2018, and the Ph.D. degree in optical engineering from Zhejiang University, Hangzhou, China. His research interests include artificial intelligence, SLAM, and 3D image processing.

DEYU GUO was born in Jining, Shandong, China, in 1999. He received the B.S. degree in physics from Zhejiang University, Hangzhou, China, in 2021, where he is currently pursuing the master's degree in electronic information. His research interests include artificial intelligence, human health monitoring, and functional near-infrared spectroscopy.

SAILING HE (Fellow, IEEE) received the Licentiate of Technology and Ph.D. degrees in electromagnetic theory from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1991 and 1992, respectively. Since then, he has been with the Royal Institute of Technology as an Assistant Professor, an Associate Professor, and a Full Professor. He is currently the Director of the Joint Research Center of Photonics (JORCEP) between KTH and Zhejiang University, China. He has first authored one monograph (Oxford University Press) and authored/coauthored over 700 articles in refereed international journals. He has given many invited/plenary talks in international conferences and has served in the leadership for many international conferences. His current research interests include subwavelength photonics, optical sensing, imaging, and communications.