Moving museums into the Metaverse

Ruier Zhang

Abstract

In recent years, the fusion of advanced technologies with virtual reality has opened new cultural preservation and engagement avenues. This dissertation explores the innovative application of Neural Radiance Fields (NeRF) technology in transcending the boundaries of physical museums and transporting their treasures into the metaverse. While classical computer vision has seen substantial progress, a developing intersection exists between NeRF and cultural heritage preservation. This study bridges this gap by introducing an approach that amalgamates NeRF techniques with the rich cultural wealth of museums.

The conventional museum experience is extended into the metaverse through a novel methodology that leverages NeRF's capabilities. The core objective is to enable individuals to explore digitized museum artifacts with unparalleled realism. NeRF technology captures intricate visual details and enables immersive interactions by rendering scenes with volumetric precision, transforming how cultural artifacts are experienced and understood.

This dissertation delves into the technical intricacies of integrating NeRF technology into the metaverse. The implementation involves the reconstruction of 3D artifact models. The results underscore the potential of NeRF to reshape the cultural heritage landscape by bridging the gap between traditional museums and the boundless possibilities of the metaverse.

Keywords: Neural Radiance Fields (NeRF), Virtual Reality, Metaverse, Cultural Heritage, Museum, Digital Preservation

1. Introduction

1.1 Background and Motivation

The Metaverse[1] is a virtual digital world consisting of a computer-generated virtual reality environment designed to simulate the real world and provide a way to interact with it. It is a comprehensive virtual space that can contain a variety of virtual worlds, virtual reality, augmented reality, and other interactive virtual experiences.

The metaverse concept originated from science fiction literature and movies and has recently gained increasing attention in the technology industry. It is seen as a new computing platform that extends the real world and can redefine how people interact with computers and people with each other.

In a metaverse, users can enter virtual environments through virtual reality devices (such as head-mounted displays, gloves, or holographic projections) to interact with other users, explore virtual geographic spaces, participate in virtual economic activities, create content, or experience virtual entertainment and social interactions. The metaverse concept is still evolving, and no unified definition exists. Different companies and organizations may understand and implement the idea of the metaverse in different ways. However, the metaverse generally represents a future vision of virtual reality that could redefine how humans interact, socialize, entertain, and work.

The Metaverse has four key features that have helped it gain acceptance and reach the masses. First, highly social----the Metaverse can transcend the limits of space and share a "physical" environment with people worldwide. This will profoundly change the way we communicate and interact with each other. The Metaverse provides a world that is a breathing, living, parallel reality that can serve all of the world's inhabitants continuously and in real time. It is hugely scalable, enabling the simultaneous coexistence of hundreds of millions of virtual characters worldwide. Second, persistence----the Metaverse will never pause or stop but will last indefinitely. The Metaverse is not limited by hardware, from computers to consoles to cell phones, and everyone can interact in the Metaverse with different types of devices. Third is interoperability- using open-source code and encryption protocols, the Metaverse can provide unprecedented interoperability of data, digital items/assets, and content in every experience. The Metaverse can make the digital world a shopping mall, where each store can use its currency with a proprietary universal ID.[2] Fourth, economic benefits----as a digital species, we will witness further transformation in the Metaverse. In the future, the Metaverse will likely be seen as a legitimate workplace and investment vehicle, offering rich content and becoming a vibrant emerging community. The Metaverse will allow users to create, invest in, own, rent, sell, or buy services as they would in the real world.[3] The future

form of the Metaverse is given to us in Sword & Sworcery, Top Gun, and Second Life. In addition, Disney's Wreck-It Ralph 2 also fits my understanding of the Metaverse: we enter the online world with a cyber doppelganger, not an off-screen player. After entering the metaverse, "I" can chat with friends, go shopping, work, or enter any game to fight; except for eating and sleeping, "I" can live in this virtual real world, which is more interesting than reality.

Here are some specific examples of elements that can be part of the metaverse:

a. Virtual Worlds: These are immersive digital environments where users can explore and interact with various objects, landscapes, and other users. Examples include games like "Fortnite" and "Minecraft," as well as virtual worlds like "Second Life."[4]

b. Social Platforms: Platforms like Facebook's Horizon Workrooms, Rec Room, and VRChat enable users to interact with others in virtual spaces, attend events, and engage in activities together.

c. Augmented Reality (AR) Applications: While the metaverse is primarily associated with virtual reality, it can also include augmented reality experiences. AR applications like Pokémon Go and Snapchat filters overlay virtual elements onto the real world, creating interactive and immersive experiences.

d. Virtual Conferences and Events: The metaverse can host virtual conferences, concerts, exhibitions, and other events where people worldwide can participate and interact with each other in a virtual environment. For example, the virtual concert series "Fortnite Presents" has featured performances by popular musicians.

These examples illustrate some diverse aspects that can be part of the metaverse, but it's important to note that the concept is still evolving. New experiences and technologies will continue to shape its development.

1.2 The Relationship between Museums and the Metaverse

With the rise of the metaverse concept, museums worldwide have started to "move" into the metaverse, transforming museum exhibits and cultural heritage into digital form and presenting them to audiences more diversely and interactively. This trend not only brings new ways of display to traditional museums but also provides a richer and more interesting cultural experience for visitors. Here are some of the highlights of metaverse displays in museums worldwide: Multi-media displays: Metaverse galleries can display static images and text and present exhibits through multiple media such as video, sound, and animation. This multi-media display allows visitors to gain a deeper understanding of the stories behind cultural relics, artworks, and other exhibits, enhancing the interactivity and interest of the cultural experience. Exhibition across time and space: The meta-universe exhibition hall can present exhibits from different eras and regions in a virtual way, allowing visitors to travel through time and space and experience the charm of different cultures

Fig.1: The Smithsonian's Arts and Industries Building (AIB) and Meta Immersive Learning will debut " FUTURES x Meta: Moonwalk"

firsthand. For example, inspired by the real-life experiences of Apollo astronauts, the Smithsonian National Air and Space Museum has created an innovative project called "Moonwalk." This project combines thousands of rare archival images, 3D scans of artifacts from the Smithsonian's collection, NASA mission audio recordings, and cutting-edge VR technology to recreate the lunar world. It allows visitors to explore previously unseen lunar landscapes. Interactive: The meta-universe exhibition hall can use virtual reality technology to allow visitors to interact with exhibits and enhance their sense of participation and experience. For instance, the world-renowned Louvre Museum offers an online guided tour, allowing users to appreciate its precious art collections remotely.

Fig.2: The photo of the homepage of the Louvre Online Museum

virtual reality technology and explore and observe the living habits of dinosaurs in virtual scenes. Social: The meta-universe exhibition hall allows for communication, sharing, and collaboration among viewers through social interaction features. For example, In the Metropolitan Museum of Art's Metaverse Gallery in New York, one can visit the exhibition with other visitors, explore cultural heritage together, and share their insights and feelings.

But there are many difficulties in promoting this project: to build a "digital community" and to make it "interactive," there is still a lack of a large interconnected platform, that is, a virtual world where all museums and their visitors can communicate and participate with each other at the same time. The portal is still lacking. In addition, the construction of the metaverse is still in the exploratory stage, with many uncertainties and immaturity, and a series of rules and systems are needed to support its normal operation.

1.3 Technology and Challenges of Digitizing Museum Exhibits

To accurately recreate the museum's exhibits and environment in the metaverse, 3D scanning and modeling are required. This can be achieved through laser scanners, photographic techniques, or depth sensors. The scanned data can create realistic 3D models for display in the metaspace. However, these methods may face the following problems in their implementation:

a. Large-scale data processing: With many exhibits and artifacts in museums, data processing and management may face large-scale data challenges. Processing and storing large amounts of 3D scan data, high-resolution images, and metadata may require huge computing and storage resources.

b. Technical requirements and equipment costs: Using laser scanners, photographic equipment, and depth sensors requires certain technical knowledge and specialized equipment. Acquiring these devices may require a high-cost investment and specialized personnel to operate and maintain the equipment.

Laser scanners use laser technology for 3D data capture and modeling. It captures a target object's geometry and surface details by emitting a laser beam and measuring the reflection time and intensity between the laser beam and the target object. The working principle of the laser scanner is based on the principle of propagation and reflection of the laser beam at the speed of light. By measuring the time difference between the laser beam and the target object, the propagation distance of the laser beam can be calculated. Combined with the position and angle information of the scanner, the 3D coordinates and geometry of the target object can be obtained. A complete 3D model of the target object can be obtained by scanning and measuring several times. Laser scanners are widely used in many fields, including engineering and construction, manufacturing, cultural heritage preservation, and digital reconstruction. It can acquire the geometry and details of the target object in a non-contact and high-precision manner, providing essential data for real-time measurement, modeling, and visualization. Laser scanners have some disadvantages, including a.Expensive: High-quality laser scanners are usually expensive, which may not be feasible for some projects with limited budgets or for personal use. b. Dependence on external conditions: Laser scanners require high environmental conditions. For example, it requires high light stability and reflectivity, so accurate scanning results may not be obtained on surfaces with unstable or low reflectivity. c. Slower scanning speed: It takes time for laser scanners to perform accurate scanning, especially for complex scenes or large objects; the scanning time may be longer. This can cause the scanning process to become time-consuming and create some limitations for real-time applications or projects that require fast results. d. Post-processing required: The raw data acquired by the laser scanner needs to be post-processed and aligned to produce an accurate 3D model. These post-processing steps may require specialized software and skills, increasing the complexity and effort of data processing. e. Inability to capture internal details: Laser scanners are mainly used to capture the target object's external geometry and surface details, while detailed information on the internal structure or details of the object is not available.

There are also several drawbacks to using photographic techniques to bring museums into the metaverse, including limitations in data acquisition: Photographic techniques rely on taking photographs to capture the museum's exhibits and environment. This means that

Each exhibit needs to be photographed individually in the real world, which can require significant time and human resources. In addition, direct photography may not be possible for some special or sensitive exhibits, resulting in incomplete or missing data. b. Limits the three-dimensional information of objects: Photography technology mainly captures two-dimensional images of objects without access to complete three-dimensional information. This means that the geometry and three-dimensionality of some objects may be lost when presented in the metaverse, limiting the user's visual experience. c. Limitations of texture and detail: Photography techniques have limitations on capturing the texture and detail of objects. The resolution of the camera and the shooting conditions may affect the quality of the image and the visibility of details. This may result in objects displayed in the metaverse lacking fine textures and details, affecting the user's realistic perception. d. The complexity of data processing and integration: Integrating and processing large amounts of photographic data into usable models and scenes requires complex data processing and computation. This may require computer vision and graphics processing techniques to process and reconstruct scenes, increasing the complexity and

time cost of data processing. e. Challenges of change and updating: Museums' exhibits are often dynamic and may have new exhibits added or old ones replaced. Data captured and presented using photographic techniques are difficult to update and reflect these changes in real-time, requiring additional data collection and processing efforts.
There are also some drawbacks to using depth sensors to move museums into the metaverse, including a. Limited range: Depth sensors typically have a limited range. Although the measurement accuracy is high at close distances, depth sensors may not provide accurate depth information at longer distances. This may lead to inaccuracies or ambiguities in the objects' depth in the metaverse. b. Sensitive to lighting conditions: The performance of the depth sensor may be affected by lighting conditions. Intense lighting or excessively dark environments may degrade the performance of the depth sensor, resulting in inaccurate depth measurements. In museums, lighting conditions may be inconsistent, especially in specific exhibition areas, which may challenge using depth sensors. c. Limitations on textured and transparent objects: Depth sensors acquire depth information primarily by projecting infrared light and measuring the reflection time. Therefore, for surfaces lacking texture or transparent objects, the performance of the depth sensor may be degraded and unable to provide accurate depth data. This may cause problems with some specific objects or exhibits in the museum. d. Data processing and noise: The raw data acquired by the depth sensor may contain noise or incomplete information. This requires data processing and filtering to remove the noise and extract accurate depth information. The complexity and computational effort of data processing depend on the sensor type and the data quality. e. Challenges for dynamic scenes: If there are dynamic scenes in the museum, such as moving people or exhibits, depth sensors may need to be able to capture and track these dynamic changes in real time. This may require a combination of other sensors or technologies to enable real-time dynamic scene capture and presentation.
The subjectivity of data annotation and description, copyright and legal issues, user experience and technological adaptability, and data security and privacy protection are key considerations when moving a museum into the metaverse. The metadata annotation and description of exhibits require expertise and may involve different interpretations from various experts. Coordinating and making trade-offs becomes essential in such cases.
Additionally, copyright and legal issues arise when digitizing exhibits, requiring permission from relevant rights holders. User experience and technological adaptability are crucial to providing an immersive and interactive experience. Ensuring the performance, stability, user-friendliness, and acceptance of virtual and augmented reality technologies poses challenges. Lastly, safeguarding valuable cultural heritage data involves encryption, access rights control, and data backup to protect data security and privacy.
Bringing museums into the metaspace requires data collection and training through 3D scanning, photography, data annotation, and description. This requires technical expertise, expert knowledge, and community engagement to ensure the museum experience in the metaverse is accurate, rich, and interactive. The tools needed to do this can be expensive and time-consuming, but we can now use NERF to bring museums to museums.

2. NeRF technology: a revolutionary approach to 3D scene modeling

Recently, many researchers have begun to explore whether the profound neural network revolution can make it possible for everyone to have the ability to capture such 3D scenes, making it as easy as taking a photograph. One innovation, sparked in 2020 by the paper "Neural Radiance Fields (NeRF) Neural Volume Rendering, an innovative technology sparked by the paper "Neural Radiance Fields (NeRF)" in 2020, has grabbed the attention of people. This new technique accepts multiple images as input to generate a compact representation of a 3D scene using a deeply fully connected network whose global can be stored in a file that is not much larger than a typical compressed image.[5] With this representation and recording method, the model can be The NeRF function obtains information about the color and density of points in a 3D space, but when a camera is used to image the scene, a pixel on the resulting 2D image corresponds to all consecutive spatial pixels on a ray from the camera. When a camera imagines the scene, a pixel on the resulting 2D image corresponds to all consecutive spatial points on a ray that starts from the camera. We need the rendering algorithm to obtain the final rendered color of this ray from all points on this ray. Rendering color from all points on this ray. NeRF is very cool and provides a way to change the ray's color from all points on the ray. NeRF is very cool and offers a new form of 3D scene modeling.

3. Data Collection and Processing Techniques

3.1 Building Your Own NeRF Dataset

However, the dataset provided by the NeRF source code is not enough to meet the research needs of many topics, so it is crucial to make your own NeRF dataset. the first step is to collect the data and then use COLMAP to obtain the camera position.

3.2 Working Principles and Steps of COLMAP

The principles of Colmap can be summarized in the following main steps:

Feature Extraction and Matching: Features are extracted from the input images, typically using algorithms like SIFT (Scale-Invariant Feature Transform) or ORB (Oriented FAST and Rotated BRIEF). Then, feature-matching algorithms, such as those based on feature descriptors, find corresponding feature points across multiple images.

Camera Pose Estimation: Using the matched feature points, Colmap can estimate the camera poses, i.e., the position and orientation of the cameras in the images.[6] This process is known as camera pose estimation or camera localization.[7]

Initial 3D Reconstruction: Based on the estimated camera poses, Colmap uses triangulation techniques to convert the matched feature points into 3D points, thus building an initial 3D reconstruction model.

Incremental Reconstruction: Colmap adopts an incremental reconstruction approach, gradually incorporating more images and feature points. New images are matched against the existing reconstruction to refine the reconstruction results further.

Graph Optimization: During the incremental reconstruction process, due to possible noise and errors in feature matching, Colmap performs global graph optimization to improve the accuracy of camera poses and 3D points.

Dense Reconstruction (Optional): If a more detailed 3D reconstruction is desired, Colmap also supports dense reconstruction. It utilizes pixel-level information between images to estimate scene depth, resulting in a denser point cloud.[8]

In summary, the principles of Colmap involve feature extraction and matching, camera pose and 3D point estimation based on matched features, incremental reconstruction with global optimization, and optional dense reconstruction for a more detailed 3D model.

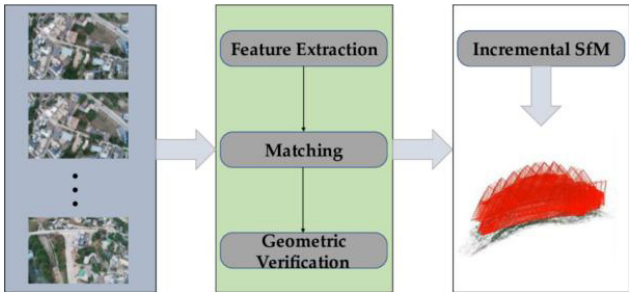

Fig.3: Principle of COLMAP[9]

3.3 Practical Museum Data Collection

I went to some museums in my area to collect data, such as Xi'an Museum, Xi'an History Museum, Xi'an Beilin Museum, etc. The artifacts I selected are more three-dimensional, such as porcelain, Buddha statues, and so on. I did not choose some flat artifacts, such as calligraphic works or murals, because these flat artifacts can be shown very well with photos. It is not necessary to use Nerf to help it realize the 3d reconstruction, and it is not as good as those more three-dimensional artifacts in the final result.

Some requirements when shooting:

a. When going to a museum, the number of images taken of an artifact should not be too small; I took about 40 images of the artifact I was focusing on trying to recreate and about 20 images of the others. I took 180 photos of 6 subjects, 168 valid photos.

b. You have to ensure a relatively obvious overlap of scenes from picture to picture.

c. The lighting situation of each picture must not be significantly different, the lighting situation of the scene must not be too bad, and the camera exposure must be locked for scenes with large lighting variations.

d. Try not to have dynamic blur; obvious dynamic blur will affect the reconstruction quality and training results.

e. The background also needs attention; if the background is not well differentiated from the object, it needs to be re-shot, and the object should account for about 90% of the whole photo.

My method is to find a good angle, keep the distance between me and the artifact, and then shoot 360 degrees around the artifact to make sure that these photos can present the complete details of the artifact. It is okay if one photo can't present all the details, you can take more photos, as long as all the photos include the details of each part of the object.

3.4 Considerations in Dataset Creation

a. The size of all the pictures must be the same. Otherwise, COLMAP cannot complete the feature extraction.

b. COLMAP reconstruction may consume many memory resources; if there is a flashback, the CPU/GPU memory is likely insufficient. You can not use high-resolution images, but the resolution is too low may fail to successfully match the location, or if the location does not allow, you should use the maximum resolution of the image as memory allows.

c. When using COLMAP to get the camera pose, I tried to format the pose data using the LLFF script and got an ERROR: Unable to access the correct camera pose for the current point. This problem can be caused by some images not matching the pose; you can try to pick out the

matching images to re-match the pose or delete the images that don't match the pose.

4. Building NeRF Network Components

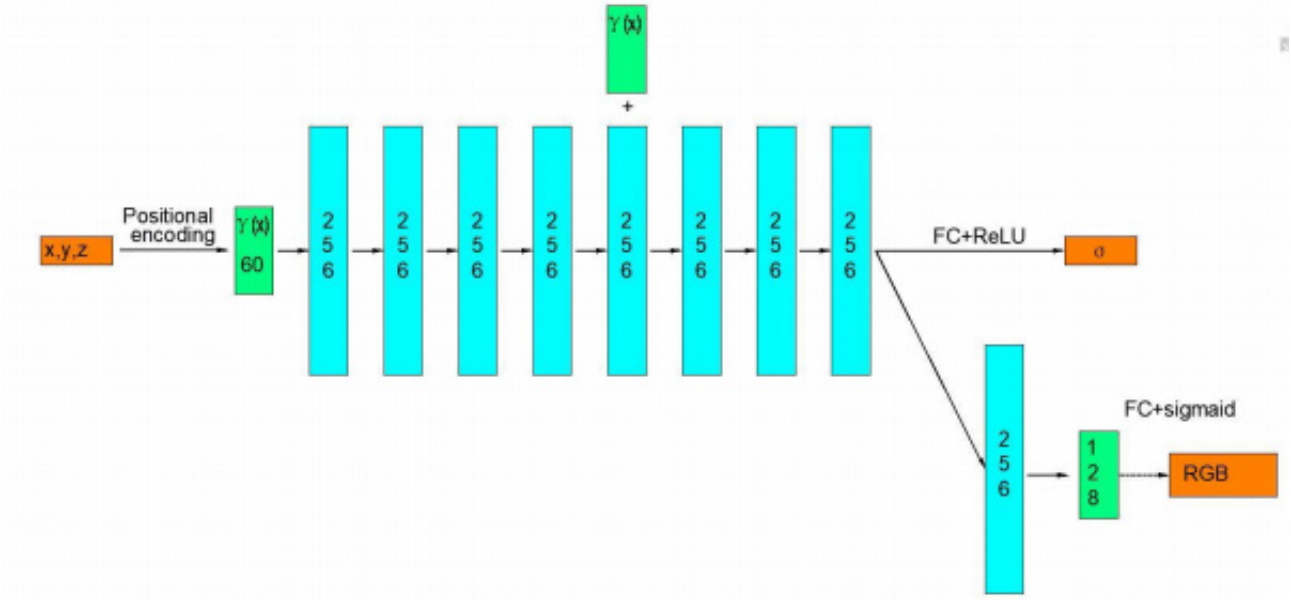

Fig.4: Network structure diagram of NERF[10]

In the NERF framework, the construction of network components is central to achieving its outstanding performance.NERF's network structures all include:[11]

a. Input Layer: Leading the Transformation from Image Coordinates to Scene Representation

The input layer of NERF plays a key role in the complexity of mapping image coordinates to models. By utilizing the image coordinates (x, y), we define a network of rays starting from the camera position. These rays cross the image plane and intersect objects in the 3D scene, paving the way for subsequent scene sampling.

In NERF, the input coordinates are considered the direction of the rays emanating from the camera center. In practical implementations, these 2D image coordinates are often fused with known camera parameters (e.g., focal length, aperture, etc.) to compute the ray's starting point and direction. This flexible input mechanism initiates converting from 2D image coordinates to 3D entities.

b. Encoder: Leading the Abstraction Journey

The encoder plays the role of the starting stage in the NERF architecture, whose task is to transform the initial 2D image coordinates into a latent, low-dimensional representation z.

This hidden representation captures the scene's local features and structural information, which provides the input for the work of the decoder.

The design of an encoder may contain multiple fully connected layers, each responsible for mapping the input features to a higher-dimensional representation. In this process, the encoder progressively extracts abstract features from the input coordinates, supporting the decoder to reconstruct the scene details better.

c. Decoder: lighting up the scene synthesis field

At the heart of the NERF network is the decoder, which receives the potential representation z generated by the encoder and reveals the brightness of each 3D scene component. The decoder converts these hidden representations into lighting, color, and geometric information about the scene.

The decoder typically consists of multiple fully connected layers that progressively map the hidden representations to higher dimensional representations to reconstruct the details of the 3D scene.[12] Each fully connected layer captures the scene information at a different level, progressively enriching the representation from global to local.

d. Perspective Sampling: Sculpting Perspectives through Synthesis

To generate images from different viewpoints, NERF employs a viewpoint sampling layer, a key link to retrieve the corresponding radiometric values from the decoder. These values are used to synthesize images based on parameters such as camera position and orientation.

The viewpoint sampling layer extracts the scene information corresponding to the new viewpoint from the output of the decoder, thus enabling the ability to view the scene from different angles. This step is crucial in NERF's computation of accurate lighting and shading effects based on different viewpoints.

By carefully coordinating the interactions between these components, the NERF network architecture can infer complex structural and appearance information about a 3D scene from simple 2D image coordinates. This comprehensive modeling capability has enabled NERF to succeed remarkably in image synthesis and scene reconstruction.

5. Experimentation and Results

5.1 NERF Model Training and Challenges

While training the NERF model, we face several problems and challenges related to the following:

a. Memory and computational requirements: The training of NERF models involves many computational operations, including forward propagation, backpropagation, and parameter updating.[13] These operations require storing information such as network

Parameters, activation values, gradients, etc., in memory and performing matrix multiplication and other operations. These computational and memory requirements increase significantly, especially when processing high-resolution images and complex 3D scenes. For NERF training with high-resolution images and complex scenes, you should use a graphics card with a large memory capacity, such as 16GB, 32GB, or higher. This ensures that the model parameters and intermediate computation results can be accommodated in memory to avoid memory overflow problems. Image resolution refers to the number of pixels in the input image. For training NERF models, it is usually recommended to use medium to high-resolution images. For example, the image resolution can be 512x512, 1024x1024, or even higher. Higher image resolution helps to capture details and the geometry of the scene. Video memory capacity is an important factor when considering the appropriate image resolution.

In cases where the GPU has 16GB of VRAM, for example, careful consideration should be given to balancing image resolution with memory limitations. This memory allocation provides a way to train NERF models on medium-sized images. Resolutions in the range of 512x512 to 1024x1024 are viable options as they allow key details and geometric features to be captured within the limits of available memory. It is worth noting that while 16GB of VRAM may impose some limitations, it is still possible to train at resolutions that provide satisfactory visual fidelity and training stability.

With the expansion to 32GB of VRAM capacity, training NERF models on high-resolution images becomes more feasible. This opens up the possibility of capturing finer details and complex scene geometry. Image resolutions from 1024x1024 to 2048x2048 can be pursued with confidence. These resolutions allow the model to capture detailed information from the scene, resulting in visually richer images.

However, even with enhanced VRAM, the choice of resolution requires careful judgment. Balancing computational resources and training efficiency is still critical, and higher resolutions may require more computational resources and training time.

b. Insufficient data: NERF requires a large amount of diverse and appropriate data for training. The model may have difficulty generalizing to new scenes or perspectives if the dataset is small or biased.

c. Perspective differences: The NERF model may generalize poorly to unseen perspectives if the training data does not cover a wide range of perspectives. This may result in artifacts or inaccuracies when rendering images from new viewpoints.

d. Depth ambiguity: NERF models may have difficulty handling scenes with depth ambiguity, such as transparent or reflective surfaces, occlusions, or scenes with significant depth variations.

e. Long Training Time: Training a NERF model can be time-consuming, especially if complex network structures or large datasets are used. Trying different hyperparameters may also increase the training time.

f. Overfitting: Given the flexibility of neural networks, NERF models may be susceptible to overfitting problems, especially when training data is limited. Regularization techniques and careful tuning of the network structure and hyperparameters can mitigate this problem.

g. Discontinuities and discrepancies: NERF may have difficulty capturing sharp discontinuities, such as object boundaries or fine geometric details. This may result in blurring or artifacts when rendering images.

h. Generalization of new scenes: NERF models are scene-specific and may have difficulty generalizing to scenes significantly different from the training data. Fine-tuning or migration learning methods may need to be used.

i. Loss function design: Designing an effective loss function is important for training NERF. A wisely chosen loss function should balance factors such as image fidelity, geometric accuracy, and regularization to ensure that the model accurately captures appearance and shape.

g. Hyper-parameter tuning: NERF has various hyper-parameters, including network architecture selection, learning rate, regularization strength, etc. The tuning of these hyper-parameters is important to obtain the best results. Tuning these hyperparameters is critical for optimal performance.

k. Rendering speed: NERF's original formulation may not be suitable for real-time rendering. It may be necessary to use multi-level representations or layered methods to

speed up rendering.
l. Complex Scenes: NERF may have difficulty with very complex scenes
with complex geometry or challenging lighting conditions. [14]

5.2 Problems and Solutions in Model Training

In our experiments, we chose a GPU equipped with 32 GB of video memory to conduct experiments on image synthesis based on the NERF model. We conducted four rounds of experiments and achieved relatively successful results in the final round. However, we also encountered some challenges and problems in the process.
In the first round of experiments, we chose a complexly textured bell as the experimental object, which had a complex texture and contained elements such as text in the background. We used 35 photos as training data, each with a resolution of 3042x4032. However, due to the high resolution of the images, the GPU with 32GB of video memory faced a memory shortage when processing these high-resolution images, which prevented the training process from proceeding properly. We adjusted the dataset by reducing the image resolution to 2000x2666 to better accommodate the graphics memory limitation. However, even in the second round of experiments, after 10,000 iterations, we still encountered the same insufficient memory problem that prevented the experiments from running smoothly.[15]
In subsequent experiments, we turned to a brightly colored drum with a relatively simple texture as our subject. Although we dropped the complex shelf texture under the drum in this
To minimize the difficulty, we resized the image and compressed it to 2000x2000. However, regrettably, even after 40,000 iterations, we encountered a loss (Loss) and a peak signal-to-noise ratio (PSNR) that became NaN.
For this situation, we can consider several potential factors. First, numerical stability may be one of the main reasons for the loss and PSNR to become NaN. Too large or small model parameters may lead to numerical instability, affecting the computational results. In deep learning, this problem can be mitigated to some extent by using numerically stable optimizers, proper parameter initialization, and regularization methods. Second, gradient vanishing or gradient explosion may lead to numerical instability, thus affecting the loss and evaluation metrics. A reasonable choice of parameters such as activation function, optimizer, and learning rate in network training and techniques such as gradient clipping can help mitigate this situation.
In addition, data anomalies may also affect the results of loss and PSNR calculation. Especially for high-resolution images, outliers or missing values may lead to computational errors, thus affecting the training process. In the data preprocessing and cleaning stages, outliers should be detected and processed to ensure the quality and stability of the data.
For optimization problems, it is also crucial to choose appropriate optimization algorithms and parameter settings. Different problems may require different combinations of optimizers and parameters to obtain better convergence and stability.
In conclusion, we encountered a series of challenges during our experiments, which included issues such as image resolution, memory constraints, and numerical stability. Although we succeeded in the last round of experiments, it reminds us that we need to consider various factors during deep learning training and make appropriate adjustments and optimizations to obtain stable and effective training results.

Fig.5: Experimental Result Graph

6. Conclusion

This study aims to explore how museums can be introduced into the metaverse through NERF technology to provide a new way of exploring culture, art, and history. Through in-depth research and practice, we have achieved remarkable results and found that NERF technology has great potential and promise in transforming museums into metaverse experiences.
First, we deeply analyzed the core components of the NERF network architecture, including encoders, decoders, and perspective sampling. The interaction of these components
Provides a viable basis for the meta-universalization of museums, which can present artifacts and exhibits in museums as high-quality 3D images to the user, enabling an immersive virtual visit experience.
Secondly, we applied NERF technology to museum meta-cosmopolitanization and successfully simulated the visual representation of museum interiors by constructing scenarios and utilizing existing digitized artifact data. This practice demonstrates the potential of NERF in cultural transmission and education, allowing people to get in touch with history and cultural heritage.
In addition, we explored the multiple applications of museum meta-universes in education, cultural exchange, and entertainment. Introducing museums into the metaverse can break down geographical and time

constraints and enable people to visit museums anytime and anywhere, thus expanding the scope and means of cultural communication. This provides a new platform for educational institutions, cultural institutions, and tourists to engage in cultural exploration in a more creative and immersive way.

However, we also note the challenges and limitations in integrating museums into the metaverse. These include aspects of data acquisition and processing, technical implementation complexity, and user experience optimization. Future research can further focus on these issues to promote the further development of museum meta-universalization.

In summary, introducing museums into the metaverse through NERF technology opens new cultural and historical transmission avenues. This innovative research provides a new direction for developing the metaverse field and positively contributes to protecting and transmitting human cultural heritage. We believe that, shortly, museum meta-universalization will become a brand-new cultural experience, bringing people a broader cultural vision and an in-depth learning experience.

References

1. O'Brian, Matt; Chan, Kelvin (October 28, 2021). "EXPLAINER: What is the metaverse and how will it work?". *ABC News*. Associated Press. Archived from the original on December 4, 2021. Retrieved December 4, 2021.

2. Chen, J. V., Farnham, S. D., & Hughes, J. J. (2011). The Metaverse as a Sociotechnical System of Systems.

3. Morgan, J. (2014). Defining the Metaverse.

4. Kakadiaris, I. A., Papadias, D., & Tzovaras, D. (2017). The Metaverse: A Collective Virtual Shared Space with Paradigms of Interaction.

5. "Computer Vision – ECCV 2020", Springer Science and Business Media LLC, 2020

6. Richard Szeliski. "Computer Vision," Springer Science and Business Media LLC, 2022

7. Schönberger, J. L., & Frahm, J. M. (2017). 3D Match: Leveraging Local Geometry for Accurate 3D Reconstruction and Matching.

8. Schönberger, J. L., Zheng, E., Pollefeys, M., & Frahm, J. M. (2018). A Benchmark for RGB-D Visual Odometry, 3D Reconstruction, and SLAM.

9. Hong, Zhonghua & Yang, Yahui & Liu, Jun & Jiang, Shenlu & Haiyan, Pan & Zhou, Ruyan & Zhang, Yun & Han, Yanling & Wang, Jing & Yang, Shuhu & Zhong, Changyue. (2022). Enhancing 3D Reconstruction Model by Deep Learning and Its Application in Building Damage Assessment after Earthquake. Applied Sciences. 12. 9790. 10.3390/app12199790.

10. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In European Conference on Computer Vision (ECCV).

11. Sitzmann, V., Zollhofer, M., & Wetzstein, G. (2021). NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In International Conference on Computer Vision (ICCV).

12. Srinivasan, P. P., Mildenhall, B., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). PlenOctrees for Real-time Rendering of Neural Radiance Fields. In ACM Transactions on Graphics (TOG), SIGGRAPH.

13. Frank Dellaert. "Neural radiance fields explode on the scene," Communications of the ACM, 2022

14. Zhou, X., Li, Z., Zhu, R., Shi, C., & Tong, X. (2020). Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. In SIGGRAPH Asia.

15. Chen, X., Shen, Y., Zhang, Y., Shen, C., & Yan, Y. (2021). SPLATNet: Sparse Lattice Networks for Point Cloud Processing. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).