3D Reconstruction Technology of Virtual Art Museum Integrating Neural Radiation Field

Xiaopeng Pei
School of Education and Arts, Hebi Polytechnic, Hebi 458030, Henan, China
Corresponding author: Xiaopeng Pei (e-mail: 15503929190@163.com)

ABSTRACT High-fidelity three-dimensional reconstruction of virtual art museums is essential for digital cultural heritage preservation and immersive visual communication, while also providing valuable references for computational imaging and electromagnetic-based scene perception under complex radiance propagation conditions. However, conventional Neural Radiance Field (NeRF) methods are limited by unstable training and texture distortion caused by specular reflections, translucent materials, and non-uniform illumination, making it difficult to preserve fine artistic details. To address these challenges, this study proposes a Distortion-Aware Ray Sampling Neural Radiance Field (DARS-NeRF) framework that integrates distortion-aware ray sampling with structure–material–lighting decoupled modeling. The proposed method enhances sampling density in visually distorted regions while combining Mask2Former semantic segmentation and a Light Estimation Network to provide semantic priors and illumination constraints for physically consistent reconstruction. Experiments conducted on 14,700 real-world art gallery images demonstrate that DARS-NeRF achieves PSNR, SSIM, and LPIPS values of 28.9 dB, 0.912, and 0.142, respectively, outperforming the strongest baseline by 2.2 dB, 3.2%, and 16.0%. In addition, Albedo MAE and Roughness RMSE are reduced by 18.3% and 20.4%, while View Consistency reaches 6.42 lux, indicating superior robustness to illumination variations. The proposed framework enables collaborative reconstruction of geometry, materials, and lighting with enhanced physical consistency, providing an effective paradigm for virtual museums, digital preservation, remote education, and high-fidelity intelligent visual sensing.

INDEX TERMS neural radiance field, distortion-aware ray sampling, virtual art museum, computational imaging, electromagnetic scene perception

I. INTRODUCTION

ARTWORKS in real art museum scenes— such as thick brushstrokes of oil paint, strong specular reflections of glass frames, highlight overflow of metal bases, and refractive distortion of translucent display cases—pose severe challenges to traditional 3D reconstruction methods [1,2]. The core challenge lies not only in the accurate restoration of geometric structures but also in the ultimate pursuit of the "artistic authenticity" of artworks: microscopic material details must be preserved losslessly, and complex lighting environments must be accurately separated and reconstructed to ensure that the reconstruction results visually meet the standards of artistic appreciation and physically support material analysis and relighting editing. Geometrydriven methods such as photogrammetry [3] and SFM-MVS (Struc-ture from Motion - Multi-View Stereo) [4,5] have difficulty capturing microscopic material details and are easily affected by non-uniform lighting; while explicit representations based on voxels or point clouds have inherent defects in continuity and texture fidelity. While NeRF (Neural Radiance Fields) can achieve high-fidelity image synthesis, its training is unstable under complex materials and dynamic lighting, and is prone to color drift, ghosting, and structural blurring [6,7]. The current lack of a reconstruction framework specifi-cally designed for cultural heritage digitization that balances physical authenticity with semantic priors se-verely restricts the largescale application of virtual art galleries in cultural heritage preservation, distance education, and immersive communication. A new implicit representation method is urgently needed that can collaboratively model geometry, materials, and lighting, and adaptively handle visual distortion.

This paper proposes the DARS (Distortion-Aware Ray Sampling)-NeRF framework, whose core technical components address three core defects in existing research: (1) To address the problem of information loss in high-light/transmittance areas caused by uniform sampling, a distortion-aware ray sampling module is designed to in-

tegrate gradients, semantic edges, and material masks to dynamically increase the sampling density at specular reflections and geometric boundaries; (2) To overcome the color drift caused by the coupled output of materials and lighting, a semantically guided ternary decoupling architecture is constructed, combining Mask2Former with the lighting estimation network to explicitly separate the output volume density, PBR material parameters, and spherical harmonic lighting; (3) To address the instability of physical decoupling caused by the lack of semantic constraints, a physical constraint loss based on semantic labels and a two-stage alternating optimization strategy are applied to ensure that the material parameters conform to the physical properties of cultural relics. To rigorously verify its effectiveness, this study evaluates three key dimensions: geometric reconstruction accuracy, material restoration fidelity, and lighting modeling consistency, comprehensively demonstrating its superiority and application value in virtual art gallery scenarios. Validated on a self-constructed dataset (consisting of 14,700 real art gallery images), DARS-NeRF outperforms competing methods in metrics such as PSNR (Peak Signal-to-Noise Ratio), Albedo MAE (Mean Absolute Error), and lighting consistency. This study applies a distortion-aware sampling mechanism for art scenes and constructs a semantically guided structure-material-lighting decoupled NeRF architecture. This framework provides a new paradigm for the precise preservation and immersive dissemination of digital cultural heritage. The term 'artistic realism' as used in this paper is a multi-dimensional evaluation concept, encompassing three aspects: geometrical accuracy, material visual fidelity, and lighting physical consistency. Specifically, it requires the reconstructed result to be geometrically faithful to the original form, accurately reproduce surface reflective properties (such as paint brushstrokes, metallic luster, and fabric texture), and maintain physically reasonable and viewpoint-consistent lighting effects. To objectively measure this goal, this paper establishes a quantitative evaluation system: geometrical accuracy is evaluated using pixel-level metrics such as PSNR and SSIM; material fidelity is measured using physical parameter errors such as Albedo MAE and Roughness RMSE; and lighting consistency is characterized by the stability and physical credibility of scene lighting through the cross-viewpoint brightness standard deviation (in lux).

II. METHOD

A. OVERALL FRAMEWORK

To achieve high-fidelity reconstruction of complex materials and non-uniform lighting in real art museum scenes, this paper proposes the DARS-NeRF framework. Its core is to build a neural rendering system that integrates semantic perception, decoupled modeling, and adaptive sampling. As shown in Figure 1, this architecture processes geometric structure, surface material, ambient lighting, and visual distortion information through parallel branches, ultimately fusing them into a unified implicit radiance field, achieving end-to-end reconstruction from sparse images to physically consistent digital twins.

The input consists of a sequence of gallery images captured from multiple perspectives and the corresponding camera parameters (intrinsic matrix, rotation matrix, and translation vector). The raw images first undergo a preprocessing pipeline to ensure data consistency and color accuracy. The preprocessed images are fed into three parallel processing branches: a semantic segmentation network, an illumination estimation network, and a DARS sampling module, providing multimodal prior information for subsequent neural field reconstruction. The semantic segmentation branch employs a Mask2Former architecture and outputs seven semantic labels, providing semantic guidance for material assignment. The illumination estimation branch uses a Convolutional Neural Network (CNN) to regress 27-dimensional spherical harmonic coefficients to characterize ambient lighting conditions. The DARS module fuses image gradients, semantic edges, and metal region masks to generate an adaptive sampling weight map for distortion-sensitive regions. The outputs of these three branches collectively provide strong prior constraints on geometry, material, and illumination for NeRF reconstruction.

The NeRF backbone receives the position-encoded 3D coordinates and viewing direction, combines the semantic labels and illumination vectors, and performs feature extraction through MLP. The network uses three independent output heads: the structure head predicts volume density and normal vectors, the material head outputs PBR parameters, and the illumination head refines the spherical harmonic coefficients. This decoupled design achieves separate modeling of geometry, material, and illumination, ensuring the interpretability and editability of physical parameters. The final output contains a continuous implicit neural field and decoupled physical parameters. The system performs end-to-end optimization through a multi-task loss function. This architecture not only supports high-quality new perspective synthesis but also outputs complete PBR material maps and environmental lighting information, providing a technical foundation for the digital preservation and immersive display of virtual art galleries.

B. DISTORTION-AWARE RAY SAMPLING

1) Distortion Source Modeling

In real art gallery scenes, the training stability and reconstruction fidelity of NeRF [8-9] are severely limited by three types of visual distortion caused by the optical properties and geometric structure of materials:

(1) Specular reflection: High-gloss surfaces such as glass picture frames and metal display cabinets produce strong directional reflection spots, leading to local pixel saturation, causing the radiation field to mistakenly encode the incident light as surface material properties, causing color drift and texture blur;

(2) Translucent transmission: Acrylic display cabinets, tulle curtains and other media have subsurface scattering and

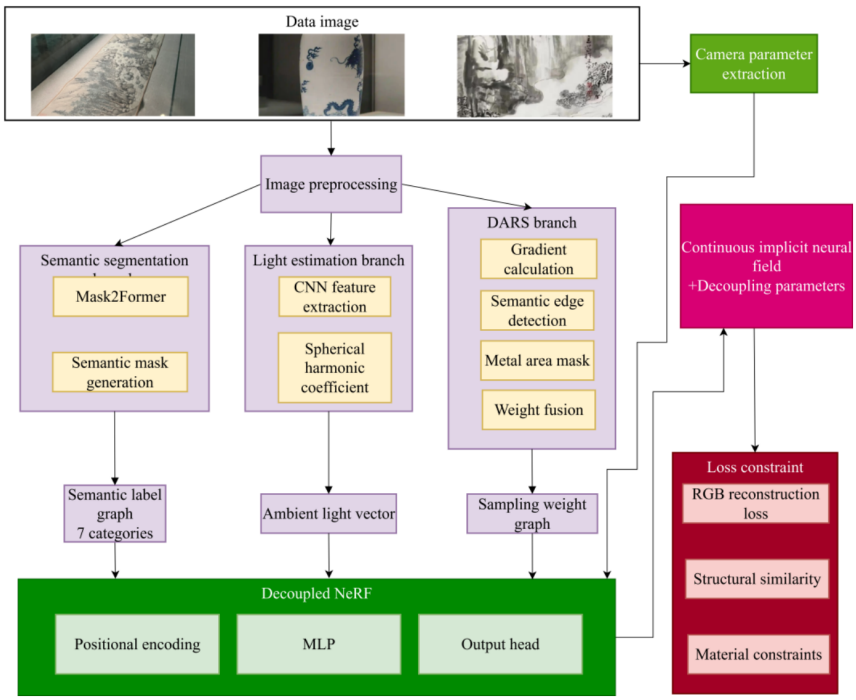

FIGURE 1. Overall Structure of this Article

refraction effects, and the light path is non-linear, destroying the "single-point ray sampling" model assumed by NeRF, resulting in depth blur and ghosting artifacts;

(3) Geometric occlusion boundary: A sharp shadow gradient is formed at the junction of the edge of the painting and the background wall. The density function in this area has highly nonlinear changes. Traditional uniform sampling cannot capture its high-frequency details, resulting in edge collapse and structural distortion.

The radiation intensity meets the following requirements:

$$L_o(x,\omega_o) = L_e(x,\omega_o) + \int_\Omega f_r(x,\omega_i,\omega_o) L_i(x,\omega_i)(\omega_i \cdot n)\, d\omega_i \tag{1}$$

In formula (1), f_r is the bidirectional reflectance distribution function, ω_i and ω_o are the incident and outgoing directions, respectively. When f_r contains a Dirac peak (specular) or a diffuse transmission component (translucency), the traditional NeRF MLP has difficulty approximating its multimodal distribution with a finite number of samples, resulting in gradient explosion or vanishing. $L_o(x,\omega_o)$ is the outgoing radiance, $L_e(x,\omega_o)$ is the self-luminous term, $L_i(x,\omega_i)$ is the incident radiance, and $(\omega_i \cdot n)$ is the cosine decay term. The 'visual distortion' described in this paper encompasses three optical phenomena with different physical mechanisms: specular reflection (nonlinear changes in local radiance in highlight regions), translucent transmission (bending of the light path caused by subsurface scattering and refraction), and geometric occlusion boundaries (high-frequency gradients at depth discontinuities). Despite their different physical causes, they all manifest as abrupt changes in local gradients or abnormal radiance distributions in the image domain, making accurate reconstruction difficult with traditional uniform sampling. The DARS module constructs a unified spatial adaptive weight map $W(x,y)$ by fusing gradient magnitude maps, semantic edge maps, and material masks. Essentially, it identifies these 'high information entropy regions' at the image feature level, rather than directly modeling the physical optical phenomena. This feature-driven fusion approach enables the sampling strategy to simultaneously respond to multiple types of difficult-to-reconstruct regions and perform non-uniform importance sampling in the light space, thereby improving the reconstruction stability and detail fidelity of complex optical interaction regions without excessively increasing the total number of sampling points.

2) DARS Algorithm

To adaptively compensate for undersampling in areas of visual distortion caused by specular reflections, translucent transmission, and geometric occlusion boundaries, this paper proposes a distortion-aware ray sampling strategy. This strategy dynamically adjusts the sampling density distribution along the ray based on semantic and gradient priors, prioritizing increased sample coverage in areas of high distortion risk. For an input image $I_i \in R^{H\times W\times 3}$, two spatially sensitive maps are computed in parallel:

Gradient Magnitude Map $G(x,y)$: This map uses the Sobel operator to extract image intensity gradients, capturing high-frequency textures and reflective boundaries.

$$G(x,y) = \|\nabla I_i(x,y)\|_2 = \sqrt{\left(\frac{\partial I_i}{\partial x}\right)^2 + \left(\frac{\partial I_i}{\partial y}\right)^2} \tag{2}$$

In formula (2), ∇I_i is a discrete gradient operator whose output range is normalized to [0, 1].

Based on the semantic mask $S(x, y) \in \{0, 1, \ldots, 6\}$ output by Mask2Former, key boundary edges such as "frame-painting" and "displaycase-background" are extracted:

$$E(x, y) = \text{Canny}\left(\text{Dilate}\left(\mathbf{1}_{S \neq \text{wall}}\right)\right) \tag{3}$$

In formula (3), $\mathbf{1}_{S \neq \text{wall}}$ is the binary mask of the non-wall area, and Dilate is a morphological dilation (kernel size = 3) used to expand the edge response range.

Distortion-sensitive weight map construction: Gradients, semantic edges, and material priors are integrated to construct a spatially adaptive weight map $W(x, y) \in [0, 1]$:

$$W(x, y) = \alpha \cdot G(x, y) + \beta \cdot E(x, y) + \gamma \cdot I_{\text{metal}}(x, y) \tag{4}$$

In formula (4), $I_{\text{metal}}(x, y)$ is the metal area mask. The output $W(x, y)$ is min-max normalized to ensure numerical stability.

Ray direction sampling density mapping: Mapping the two-dimensional weight map $W(x, y)$ to the three-dimensional ray space. For any ray $r(t) = o + td(t \in [t_n, t_f])$, its sampling density function along the depth t is defined as:

$$p(t) \propto W(\Pi(r(t))) \cdot p_0(t) \tag{5}$$

In formula (5), $\Pi(\cdot)$ is the projection function that projects the 3D point $r(t)$ to the image plane coordinates. p_0 is the baseline stratified sampling distribution, defined as:

$$p_0(t) = \frac{1}{N_b} \sum_{i=1}^{N_b} U(t_{i-1}, t_i), \quad t_i = t_n + \frac{i}{N_b}(t_f - t_n) \tag{6}$$

In formula (6), N_b is the number of stratified intervals and U is the uniform distribution.

Importance sampling and threshold truncation: Using inverse transform sampling to generate N_s sampling points $\{t_k\}_{k=1}^{N_s}$ from $p(t)$. Calculating the cumulative distribution function:

$$F(t) = \int_{t_n}^{t} p(\tau)\, d\tau \tag{7}$$

A threshold cutoff mechanism is applied: a sampling point is retained only when $W(\Pi(r(t))) > \tau$; otherwise, it is discarded and resampled ($\tau = 0.7$).

To prevent sampling instability caused by wild fluctuations in $W(x, y)$ in local regions, a smoothing regularization term is applied:

$$L_{\text{distortion}} = \lambda_{\text{reg}} \cdot \left(\|\nabla_x W\|_2^2 + \|\nabla_y W\|_2^2\right) \tag{8}$$

In formula (8), $\lambda_{\text{reg}} = 0.3$, which penalizes the spatial gradient of the weight map and improves the smoothness of the sampling distribution and training stability.

The DARS sampling effect is shown in Figure 2.

Standard NeRF allocates only three sampling points in the highlight area, which cannot fully capture highfrequency optical effects such as specular reflection and refraction, resulting in color drift and texture distortion in the reconstruction. DARS-NeRF constructs a distortion-sensitive weight map by fusing image gradients, semantic edges, and material priors, and dynamically adjusts the ray sampling density distribution. It increases the number of sampling points in the specular reflection area (highlight area) to 12, improving the ability to fit the multimodal distribution of the bidirectional reflectance distribution function; In shadow regions where the illumination intensity is significantly below the threshold and the gradient change is gradual, the DARS algorithm reduces the number of sampling points to 1. This strategy is based on the prior judgment that the light contribution in such regions is extremely low and the depth change is gradual, aiming to avoid redundant sampling in information-sparse regions. At the same time, through the constraints of neighborhood sampling interpolation and geometric regularization terms (such as Eikonal loss), the system can still maintain the geometric continuity and edge integrity of the region, thereby improving sampling efficiency while avoiding edge collapse or depth artifacts. This adaptive sampling strategy effectively compensates for the visual distortion caused by the optical properties of complex materials, improving the reconstruction accuracy of highly reflective areas, optimizing overall training efficiency, and achieving optimal allocation of sampling resources.

C. STRUCTURE-MATERIAL-ILLUMINATION DECOUPLED NERF ARCHITECTURE

1) Multi-Branch NeRF Design

To achieve explicit decoupled modeling of the geometric structure, surface material, and ambient lighting in art gallery scenes, this paper constructs a multi-head neural radiance field architecture guided by both semantics and lighting conditions. This architecture, based on a standard NeRF MLP backbone [10,11], integrates a Mask2Former semantic segmentation network and a Light Estimation Network, providing pixel-level semantic priors and global illumination vectors, respectively, as conditional inputs to guide the physically consistent outputs of the material and lighting branches. The overall architecture consists of a shared hidden layer backbone and three independent output heads, which predict volume density, PBR material parameters, and ambient lighting components, respectively.

The backbone is a multi-layer perceptron with positional encoding, taking as input the 3D spatial coordinate $x \in R^3$ and the viewing direction $d \in S^2$. Positional encoding enhances high-frequency representations:

$$\gamma(v) = \left[\sin(2^0 \pi v), \cos(2^0 \pi v), \ldots, \sin(2^{L-1} \pi v), \cos(2^{L-1} \pi v)\right] \tag{9}$$

In formula (9), v is the input vector.

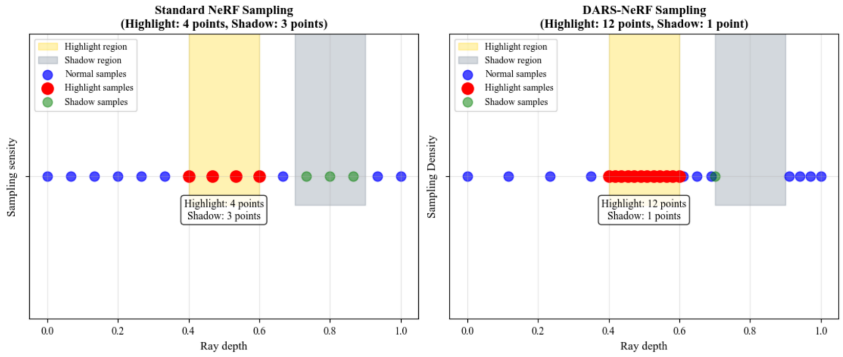

FIGURE 2. DARS Sampling Effect

The spatial coordinate x is set to 10 and the direction d is set to 4 to balance the modeling of geometric details and view correlation. The backbone MLP structure has 8 layers, each with 256 dimensions, using the ReLU (Rectified Linear Unit) activation function and LayerNorm normalization:

$$h = MLP_{\text{backbone}}\left([\gamma(x); \gamma(d); s; l_{\text{env}}]\right) \tag{10}$$

In formula (10), s is the one-hot encoding of the semantic category (from Mask2Former), and l_{env} is the ambient light spherical harmonic coefficient (from the Light Estimation Network). These two serve as conditional inputs to guide the network to output reasonable physical properties in different regions. Mask2Former [12,13] provides pixel-level semantic masks to guide the material header to output parameters that conform to physical priors in different regions and identify high-distortion boundaries.

2) Lighting Estimation Module

The Light Estimation Network explicitly estimates the spherical harmonic coefficient l_{env} of ambient lighting, decoupling lighting from materials and supporting post-processing relighting and cross-view consistency optimization.

In the Light Estimation Network, the input is an image $I \in R^{H\times W\times 3}$, which is extracted using ResNet-34 to extract global features. A three-layer fully connected network ($512 \to 256 \to 27$) outputs a 27-dimensional vector (9th-order spherical harmonic coefficients $\times$ 3 channels). This is then mapped to a reasonable physical range using sigmoids.

$$l_{\text{env}} = FC_{\text{light}}\left(\text{GlobalAvgPool}(\text{ResNet34}(I))\right) \tag{11}$$

The spherical harmonic coefficients can reconstruct the ambient light probe $L_{\text{env}}(\omega)$:

$$L_{\text{env}}(\omega) = \sum_{l=0}^{2}\sum_{m=-l}^{l} l_{\text{env}}^{(l,m)} Y_l^m(\omega) \tag{12}$$

In formula (12), Y_l^m is the spherical harmonic basis function, which is used for the ambient light integral term in the rendering formula.

Three independent output headers are used: the structure header outputs volume density to control surface visibility; the material header outputs PBR parameters [14-15] to support physically realistic rendering and material editing; and the lighting header outputs ambient lighting.

The formula for the structure header is:

$$\sigma = MLP_{\sigma}(h) = \text{softplus}(W_{\sigma}h + b_{\sigma}) \in R^{+} \tag{13}$$

For volume rendering transmittance calculation:

$$T(t) = \exp\left(-\int_{t_n}^{t} \sigma(r(s))\, ds\right) \tag{14}$$

The material header outputs PBR parameters [16-17], and the formula for Albedo (diffuse base color) is:

$$a = \sigma_{\text{rgb}}\left(MLP_{\text{albedo}}(h)\right) \in [0,1]^3 \tag{15}$$

In formula (15), σ_{rgb} is Sigmoid, ensuring that the color value is legal.

Roughness:

$$r = \sigma\left(MLP_{\text{roughness}}(h)\right) \in [0,1] \tag{16}$$

Metallic:

$$m = \sigma\left(MLP_{\text{metallic}}(h)\right) \in [0,1] \tag{17}$$

Normal:

$$n = \frac{MLP_{\text{normal}}(h)}{\|MLP_{\text{normal}}(h)\|_2} \in S^2 \tag{18}$$

In the lighting header, l_{env} is predicted by an independent network, but to support end-to-end fine-tuning, the RGB values output by the material header still need to be coupled with the lighting rendering:

$$c_{\text{render}} = a \cdot L_{\text{diffuse}} + f_{\text{spec}} \cdot L_{\text{specular}} \tag{19}$$

In formula (19), the diffuse reflection term L_{diffuse} and the specular term L_{specular} are calculated by l_{env} and parameters (using the Cook-Torrance model) and are ultimately used for pixel loss optimization.

3) Material Constraint Loss and Lighting Decoupling Training Strategy

To prevent the network from outputting material parameters inconsistent with physical priors when lacking strong supervision, this paper applies a semantically guided physical constraint loss L_{mat}. During training, soft constraints are imposed on three parameters: albedo, roughness, and metallic, ensuring they conform to the typical physical properties of cultural relic materials.

The material constraint loss is:

$$L_{\text{mat}} = \lambda_1 \cdot \|a - a_{\text{gt}}\|_2 + \lambda_2 \cdot \|r - 0.8 \cdot I_{\text{paint}}\|_2 + \lambda_3 \cdot \|m - 0.9 \cdot I_{\text{metal}}\|_2 \quad (20)$$

In formula (20), a_{gt} is the true albedo. r is the predicted roughness, and m is the predicted metalness. I_{paint} is the binary mask of the "painting" class, and I_{metal} is the mask of the metal subclass in "frame" or "base".

To stabilize the training process and avoid optimization oscillations caused by the coupling of material and lighting parameters, this paper adopts a two-stage alternating optimization strategy. First, the lighting estimation network is fixed and the geometry and material branches are pretrained. Then, the lighting and NeRF parameters are updated alternately to achieve progressive decoupling.

Freezing the Light Estimation Network parameters and optimizing only the NeRF backbone and material header, using the standard NeRF loss:

$$L_{\text{pretrain}} = L_{\text{rgb}} + L_{\text{ssim}} + L_{\text{mat}} \quad (21)$$

Illumination-NeRF alternating optimization, fix NeRF parameters, optimize illumination l_{env}, minimize reprojection error, so that the predicted illumination can explain the observed image:

$$\min_{l_{\text{env}}} L_{\text{reproj}} = \|I_{\text{pred}}(l_{\text{env}}) - I_{\text{gt}}\|_1 \quad (22)$$

In formula (22), I_{pred} is the image rendered using the current NeRF parameters and l_{env}.

Fix l_{env} and optimize NeRF parameters: Minimize pixel-level and perceptual loss:

$$\min_{\theta} L_{\text{rgb}} + \lambda_1 L_{\text{ssim}} + \lambda_2 L_{\text{psnr}} + \lambda_3 L_{\text{mat}} + \lambda_4 L_{\text{eikonal}} \quad (23)$$

The illumination estimation network takes a single-view image as input, extracts global visual features using ResNet-34, and regresses 27-dimensional spherical harmonic coefficients to represent ambient illumination. Although a single view cannot directly observe the entire environment, the network, by learning from a large number of prior distributions of the art gallery scene, can reasonably infer the intensity, direction, and color distribution of the overall illumination from cues such as material reflection, shadow distribution, and specular morphology in the visible area. Furthermore, this initial estimation is co-optimized with the multi-view rendering results in subsequent alternating optimization stages: the NeRF branch is optimized by fixing the illumination parameters, and then the illumination coefficients are optimized by fixing the NeRF parameters again to minimize multi-view reprojection errors, thereby gradually converging the illumination estimation to a globally consistent solution that can explain all training views.

D. LOSS FUNCTION DESIGN

The total loss function L_{total} in this paper takes into account pixel fidelity, structural similarity, material physics, geometric regularity, and sampling stability:

$$L_{\text{total}} = L_{\text{rgb}} + \lambda_1 \cdot L_{\text{ssim}} + \lambda_2 \cdot L_{\text{psnr}} + \lambda_3 \cdot L_{\text{mat}} + \lambda_4 \cdot L_{\text{eikonal}} + \lambda_5 \cdot L_{\text{distortion}} \quad (24)$$

In formula (24), L_{rgb} is the L1 pixel loss; L_{ssim} is the structural similarity loss; L_{psnr} is the PSNR maximization; L_{mat} is the material physical constraint; L_{eikonal} is the geometric regularization term; and $L_{\text{distortion}}$ is the DARS weight regularization.

$$L_{\text{rgb}} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{C}_i - C_i^{\text{gt}}\|_1 \quad (25)$$

In formula (25), $\hat{C}_i$ is the predicted pixel color; C_i^{gt} is the actual image pixel; and N is the number of rays in the batch.

The formula for L_{ssim} is:

$$L_{\text{ssim}} = 1 - \text{SSIM}(\hat{I}, I_{\text{gt}}) \quad (26)$$

In formula (26), a 11×11 Gaussian window is used to calculate the local structural similarity. The formula for L_{psnr} is:

$$L_{\text{psnr}} = -10 \cdot \log_{10} \left(\frac{1}{N} \sum_{i=1}^{N} \|\hat{C}_i - C_i^{\text{gt}}\|_2^2 \right) \quad (27)$$

The formula for L_{eikonal} is:

$$L_{\text{eikonal}} = E_{x \sim V} \left[\left(\|\nabla_x \sigma(x)\|_2 - 1 \right)^2 \right] \quad (28)$$

The formula for $L_{\text{distortion}}$ is:

$$L_{\text{distortion}} = \|\nabla_x W\|_2^2 + \|\nabla_y W\|_2^2 \quad (29)$$

In formula (29), a gradient penalty is applied to the distortion-aware weight map W to prevent local oscillations. The hyperparameter configuration is shown in Table 1.

TABLE 1. Hyperparameter Configuration

Loss item	Weight	Effect	Activation phase
L_{rgb}	1	Pixel matching	Entire process
L_{ssim}	0.2	Structure preserving	Entire process
L_{psnr}	0.1	Objective indicator optimization	Entire process
L_{mat}	0.5	Material physical constraints	Entire process
L_{eikonal}	0.01	Geometric smoothness	Step $>$ 2000
$L_{\text{distortion}}$	0.3	Sample weight smoothing	Entire process

III. EXPERIMENTS

A. DATASET CONSTRUCTION

In shadow regions where the illumination intensity is significantly below the threshold and the gradient change is gradual, the DARS algorithm reduces the number of sampling points to 1. This strategy is based on the prior judgment that the light contribution in such regions is extremely low and the depth change is gradual, aiming to avoid redundant sampling in information-sparse regions. At the same time, through the constraints of neighborhood sampling interpolation and geometric regularization terms (such as Eikonal loss), the system can still maintain the geometric continuity and edge integrity of the region, thereby improving sampling efficiency while avoiding edge collapse or depth artifacts. Information on these art exhibition areas is shown in Table 2.

The artistic image is shown in Figure 3.

To ensure the authenticity and accessibility of the reconstructed data, professional-grade equipment is used to capture high-precision, multi-viewpoint images of 10 representative art exhibition areas under closed-door conditions. The primary capture device is a Sony A7R V full-frame mirrorless camera equipped with a 61- megapixel sensor, with fixed settings of ISO (International Organization for Standardization) 100, shutter speed 1/125 second, and aperture f/8 to minimize noise and maintain a wide depth of field. The lens used is a Sony FE 24-70mm f/2.8 GM II, with a locked focal length of 35mm, to approximate the human eye's perspective, avoid wide-angle distortion, and ensure consistent spatial proportions.

The shooting platform used a Kessler Crane CineSlider motorized gimbal, supporting a theoretical sampling capability of 360° horizontally in 1° increments and ±30° vertically in 0.5° increments. If shooting at this density for full coverage, the theoretical maximum number of frames per exhibition area could reach 26,280 (360×73). However, in actual acquisition, limited by closing time, storage capacity, and viewpoint redundancy, we did not perform full-density acquisition. Instead, based on the exhibition area layout and visual importance, we selectively sampled from key viewpoint intervals, ultimately retaining 120 to 180 high-quality valid frames for each exhibition area for training and testing. The theoretical frame count is only used to illustrate the device's accuracy and sampling upper limit. The actual dataset totaled 14,700 images (approximately 2.8 TB), covering 10 exhibition areas, averaging 1,470 images per area. The number of valid frames used for single-scene modeling is within the feasible range for actual acquisition and reconstruction. All capture is automatically controlled by a Dell Precision 7760 workstation running a Python script, synchronizing the entire process from gimbal movement, shutter triggering, image upload, and metadata recording (including timestamps, pose, and exposure parameters). To ensure color accuracy, an X-Rite ColorChecker Passport Photo 2 standard color chart is placed in the lower left corner of the image (at least 1.5 meters from the exhibits) before each capture. Three exposures (-2EV, 0EV, and +2EV) are used for post-process white balance correction and color space restoration. A Konica Minolta CL-200A photometer is used to measure the illuminance (lux) and color temperature (K) at the center of the booth, which served as the basis for the lighting complexity annotation.

To obtain the ground truth data required for the assessment, small, independent exhibits are scanned with an Artec Leo handheld 3D scanner at 0.1mm accuracy. Artec Studio software is used to generate mesh models and 4K resolution Albedo, Normal, Roughness, and Metallic texture maps. For large, flat exhibits, a Specim FX17 multispectral imaging system is used to collect 12-band spectral data, and a physical inversion algorithm is used to generate material parameter maps. Ground truth ambient lighting values are obtained using an Insta360 Pro 2 360° panoramic camera, capturing three-exposure HDR (High Dynamic Range) sequences from the same camera position. These images are merged into 2048×1024 HDR images using Adobe Lightroom. The Light Probe Image Analyzer tool is then used to extract the ninth-order spherical harmonic coefficients (a total of 27 dimensions) as a baseline for the lighting assessment. All data collection processes strictly adhere to cultural relic preservation regulations, prohibiting the use of flash photography and contact with exhibits to ensure non-invasive and ethical data acquisition.

Based on the classification of cultural relics and physical materials, the Mask2Former annotation standard has 7 categories, and the category definitions are shown in Table 3.

Semantic segmentation and annotation are performed using the open-source labeling platform Label Studio v1.10. The annotation task is undertaken by three master's students with backgrounds in cultural heritage digitization. All personnel undergo two weeks of specialized training covering cultural relic material classification standards, edge determination criteria, and annotation tool operation specifications. To ensure annotation quality, a two-person independent annotation system is implemented, with each image independently completed by two annotators. The system automatically calculates the Intersection over Union (IoU) value. If the IoU value is $>$ 0.92, the image is accepted. If it is below the threshold, the image is submitted to a doctoral supervisor with at least five years of experience in digital cultural heritage research for review. The final annotation

TABLE 2. Art Exhibition Area Information

Serial number	Core exhibit types	Lighting conditions
M01	Song Dynasty ink scroll, Ming and Qing Dynasty meticulous painting	Natural light+spotlight mixture
M02	Shang and Zhou bronze tripods, titles, and gongs	Strong directional spotlight+mirror reflection
M03	Qing Dynasty glass ornaments, European crystal vases	Omnidirectional LED (Light Emitting Diode)+spotlight
M04	Van Gogh's "Starry Night" imitation, Monet's "Water Lilies"	Top linear light+local tracking light
M05	Song Dynasty Ru kiln sky blue glazed bowl, Ming Dynasty blue and white porcelain	Diffuse reflection+specular highlight
M06	Ming Dynasty wood carved Buddha statues and Han Dynasty pottery figurines	Low illumination+side lighting
M07	Mirror stainless steel sculpture, LED light box	Dynamic light source+multiple reflections
M08	Rare ancient books, silk albums, silk	Soft diffuse light
M09	Qing Dynasty Dian Cui Feng Guan and Diamond Brooch	Extremely strong point light source+occlusion
M10	Table, Four Treasures of the Study, Screen	Natural window light+soft light

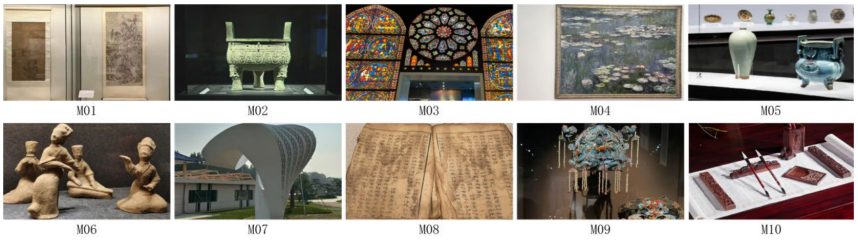

FIGURE 3. Artistic Image Display

TABLE 3. Semantic Segmentation Annotation Categories

Number	Category name	Description	Example
1	painting	Artworks such as oil painting, ink painting, printmaking, etc	Van Gogh's "Starry Night" Canvas Area
2	frame	Picture frame, picture frame, wooden/metal frame	Golden lacquer carved painting frame, aluminum modern frame
3	display	Glass/acrylic display cabinets, sealed display boxes	Transparent glass cabinet, low reflection coating cabinet
4	base	Base, booth, stone pier, metal bracket	Bronze stone pedestal, sculpture brass base
5	wall	Exhibition hall walls and background panels	Off white gypsum wall, deep red velvet wall
6	floor	flooring	Marble, wooden flooring, carpet
7	other	Unrelated objects (signs, lighting fixtures, fire hydrants)	Guide signs, spotlights, safety cameras

results are reviewed and confirmed by the supervisor before being stored.

The semantic category settings include non-artwork categories such as 'walls', 'floors', and 'others', primarily based on the following considerations: First, although these categories are not the main subjects of art appreciation, their material properties (such as the reflectivity of marble floors and the diffuse reflection of wall paint) still affect the overall consistency of scene lighting. Accurate material allocation helps improve the physical realism of global lighting decoupling. Second, the DARS sampling mechanism relies on semantic edge maps to identify geometric occlusion boundaries. Boundaries such as 'wall-frame' and 'floor-base' are also areas prone to high-frequency distortion, requiring semantic guidance to enhance sampling. Furthermore, objects such as fire hydrants and indicator lights included in the 'others' category, although not artistic components, may produce specular reflections on their metal or glass surfaces. Insufficient sampling can lead to local artifacts, thus affecting overall visual coherence. Therefore, a complete semantic system not only serves the material reproduction of the artwork itself but also provides the necessary structural priors for light-material decoupling and adaptive sampling across the entire scene.

Dataset statistics are shown in Table 4.

TABLE 4. Dataset Statistics

Statistical item	Value	Statistical item	Value
Total number of exhibition areas	10	Average illuminance (lux)	168 ± 42
Total number of images	14,700	Average color temperature (K)	4420 ± 310
Average number of images per exhibition area	1,470	Number of personnel marked	3
Number of semantic categories	7	Per capita annotation time (h)	480
Number of exhibition areas with real HDRI values	10	Annotate IoU	0.951
Dataset size (TB)	2.8	kappa coefficient	0.937

Note: HDRI (High Dynamic Range Image).

B. COMPARISON METHODS AND IMPLEMENTATION DETAILS

To comprehensively evaluate the performance of this method, it selects the most representative neural rendering methods as benchmarks, including Standard NeRF, Mip-

NeRF 360, InstantNGP, LERF, VolSDF, Hyper-NeRF, D-NeRF, and the proposed DARS-NeRF. All methods are replicated in a unified experimental environment using the PyTorch framework, employing an 8-layer, 256-dimensional MLP backbone network with ReLU activation functions and LayerNorm. Each ray sampled 64 points during training and 128 points during testing. The optimizer used is AdamW (learning rate = 5×10^{-14}, weight_decay = 1×10^{-14}, batch_size = 1024) to ensure a fair comparison.

Although D-NeRF was initially designed for dynamic scenes, its framework of implicit field modeling under time conditions demonstrates strong representation capabilities for static complex materials and lighting modeling as well. In our experiments, we set the temporal dimension of D-NeRF to zero, retaining only its multi-branch implicit field architecture to make it applicable to static scene reconstruction. Compared to other static NeRF variants, D-NeRF's architecture is closer to the structure-material-lighting decoupling design proposed in this paper, and its performance has been validated as a strong benchmark in several static complex scene reconstruction works. Therefore, choosing D-NeRF as the primary comparison method allows for a fairer evaluation of the performance gains of our proposed decoupling and sampling modules on a similar architecture, rather than directly comparing the merits of dynamic and static methods.

All models are trained on an NVIDIA A100 80GB GPU; the inference phase is deployed on a Meta Quest 3 VR (Virtual Reality) headset and in a browser to verify real-world performance. The experimental environment maintains fixed lighting conditions, camera pose, and training viewpoint, with the only variables being the model architecture and sampling strategy, ensuring that the evaluation results reflected only algorithmic differences. This setup not only addresses core aspects such as material decoupling, lighting modeling, dynamic adaptation, and geometric stability but also considers real-time performance and deployment feasibility, fully validating the comprehensive advantages of DARS-NeRF in a real-world art museum setting.

C. EVALUATION METRICS

To comprehensively, objectively, and reproducibly evaluate the reconstruction performance of DARS-NeRF in virtual art gallery scenes, this study constructs a quantitative evaluation system covering three dimensions: geometric accuracy, material fidelity, and lighting consistency.

Geometric accuracy measures the fidelity between the reconstructed image and the ground truth at both the pixel and perceptual levels.

PSNR [18,19] measures the mean squared error between the reconstructed image and the ground truth image and is defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{30}$$

$$MSE = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\|\hat{I}(i,j) - I_{\text{gt}}(i,j)\|_2^2 \tag{31}$$

In formulas (30) and (31), $MAX_I = 1.0$ (the normalized RGB value range), and H and W are the image height and width.

SSIM (Structural Similarity Index Measure) measures the consistency of local structure, brightness, and contrast, and is calculated using an 11×11 Gaussian window:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{32}$$

In formula (32), μ_x and μ_y are local means; σ_x^2 and σ_y^2 are variances; and σ_{xy} is covariance. LPIPS (Learned Perceptual Image Patch Similarity) measures differences in human visual perception based on the feature distance of the AlexNet intermediate layer:

$$LPIPS(\hat{I}, I_{\text{gt}}) = \sum_l \frac{1}{H_l W_l}\sum_{h,w} w_l \|\phi_l(\hat{I})_{h,w} - \phi_l(I_{\text{gt}})_{h,w}\|_2^2 \tag{33}$$

In formula (33), ϕ_l is the feature map of the l-th layer, and W_l is the pre-trained weight.

Material restoration is used to evaluate the closeness of the predicted PBR material parameters to the physical truth. It is calculated only on 18 exhibits with scanned truth values:

Albedo MAE measures the deviation of diffuse color prediction:

$$\text{Albedo MAE} = \frac{1}{N}\sum_{i=1}^{N}\|\hat{a}_i - a_{\text{gt},i}\|_1 \tag{34}$$

In formula (34), $\hat{a}_i$ is the predicted albedo and $a_{\text{gt},i}$ is the true value.

Roughness RMSE (Root Mean Square Error) measures the accuracy of surface roughness prediction:

$$\text{Roughness RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{r}_i - r_{\text{gt},i})^2} \tag{35}$$

In formula (35), $\hat{r}_i$ and $r_{\text{gt},i}$ are the predicted value and the true value, respectively.

Lighting consistency is used to evaluate the physical plausibility and cross-view stability of environmental lighting modeling. Env Light L2 measures the difference between the spherical harmonic coefficients extracted from the predicted lighting and the true HDRI:

$$\text{Env Light L2} = \|\hat{l}_{\text{env}} - l_{\text{env}}^{\text{gt}}\|_2 \tag{36}$$

In formula (36), $\hat{l}_{\text{env}}$ is the predicted value, and $l_{\text{env}}^{\text{gt}}$ is extracted from the real HDRI by the Light Probe Image Analyzer.

View Consistency measures the standard deviation of the brightness of the same object rendered at different viewing angles, reflecting the robustness of the lighting model:

$$\text{View Consistency} = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(I_k - \bar{I})^2} \tag{37}$$

$$\bar{I} = \frac{1}{K}\sum_{k=1}^{K} I_k \tag{38}$$

In formulas (37) and (38), I_k is the average brightness of the same pixel area at the k-th viewing angle, and $K = 10$ is the number of uniformly sampled viewing angles.

IV. RESULTS AND ANALYSIS

A. COMPARISON OF QUANTITATIVE RESULTS

The quantitative comparison results of geometric accuracy are shown in Table 5.

TABLE 5. Quantitative Comparison Results of Geometric Accuracy

Method	PSNR (dB)	SSIM	LPIPS
NeRF	24.3	0.843	0.218
Mip-NeRF 360	26.1	0.872	0.185
InstantNGP	25.8	0.861	0.192
LERF	26.4	0.879	0.176
VolSDF	25.2	0.851	0.201
HyperNeRF	25.6	0.857	0.195
D-NeRF	26.7	0.884	0.169
DARS-NeRF	28.9	0.912	0.142

DARS-NeRF outperforms all compared methods in three key metrics: PSNR, SSIM, and LPIPS. Specifically, DARS-NeRF achieves a PSNR of 28.9 dB, a 2.2 dB improvement over the best baseline, D-NeRF; SSIM reaches 0.912, a 3.2% improvement over D-NeRF; and LPIPS drops to 0.142, a 16.0% decrease over D-NeRF. This performance improvement is primarily attributed to three core innovations: First, the distortion-aware ray sampling strategy significantly increases the number of effective samples in specular and semi-transparent areas by dynamically adjusting the sampling density, avoiding the undersampling problem of traditional uniform sampling in high-frequency detail areas; second, the semantically guided multi-branch architecture explicitly decouples geometry, material, and lighting components, eliminating coupling errors between different physical components; and third, the physical constraint loss function ensures the rationality of material parameters and prevents parameter combinations that do not conform to physical priors. DARS-NeRF demonstrates greater robustness in scenes where traditional NeRF often fails, such as those with high reflectivity and complex transmissive media. This is directly reflected in the improved LPIPS metric, indicating that its reconstruction results are closer to real-world observations in terms of perceptual quality. These results demonstrate that a neural rendering framework that integrates semantic priors with physical constraints can effectively address reconstruction distortion issues caused by complex materials and non-uniform lighting in art gallery scenes.

The significance statistics of the results are shown in Figure 4.

DARS-NeRF significantly outperforms all compared methods in albedo reconstruction accuracy, reducing the mean to 5.89×10^{-2}, an 18.3% reduction compared to the best baseline, D-NeRF. This advantage stems primarily from its semantically guided material decoupling architecture and physically constrained loss function. The pixel-level semantic prior provided by Mask2Former ensures that albedo predictions for regions of different materials conform to their physical properties, while the albedo constraint in the loss directly optimizes the accuracy of diffuse color. Traditional methods, particularly on artworks with complex textures, experience color drift due to the coupling of lighting and material components. DARS-NeRF eliminates this error through explicit decoupling, achieving more accurate reconstruction of the material's true color.

In terms of surface roughness prediction, DARS-NeRF demonstrates significant accuracy improvements (RMSE = 8.26×10^{-2}), a 20.4% improvement over D-NeRF. This is primarily attributed to the physical prior constraints for different semantic categories in its material header. The DARS module's dense sampling of brushstroke textures and highlight boundary regions provides richer surface detail information, enabling the MLP to better learn the spatial distribution of roughness and avoid the smoothing of high-frequency details caused by insufficient sampling in traditional methods.

DARS-NeRF achieves a significant advantage in ambient lighting estimation accuracy (Env Light L2 = 10.57×10^{-2}). This improvement stems from its independent lighting estimation network and alternating optimization strategy. The lighting estimation network explicitly models ambient lighting by regressing spherical harmonic coefficients, avoiding the problem of traditional methods implicitly encoding lighting information into material parameters. The two-stage training strategy first pre-trains the geometry and material branches with fixed lighting parameters, followed by alternating optimization to ensure the decoupling of lighting from materials. This design enables the network to accurately separate the lighting components of a scene, particularly in the non-uniformly illuminated art gallery environment, effectively estimating complex lighting distributions and providing an accurate physical basis for subsequent re-lighting applications.

DARS-NeRF achieves the best performance in terms of cross-view consistency (View Consistency = 6.42 lux). This advantage is primarily due to its complete lighting-material decoupling architecture and physically realistic rendering formulas. By modeling ambient lighting separately from surface materials, DARS-NeRF ensures that the rendering results of the same surface at different viewpoints conform to physical laws, avoiding view-related deviations caused

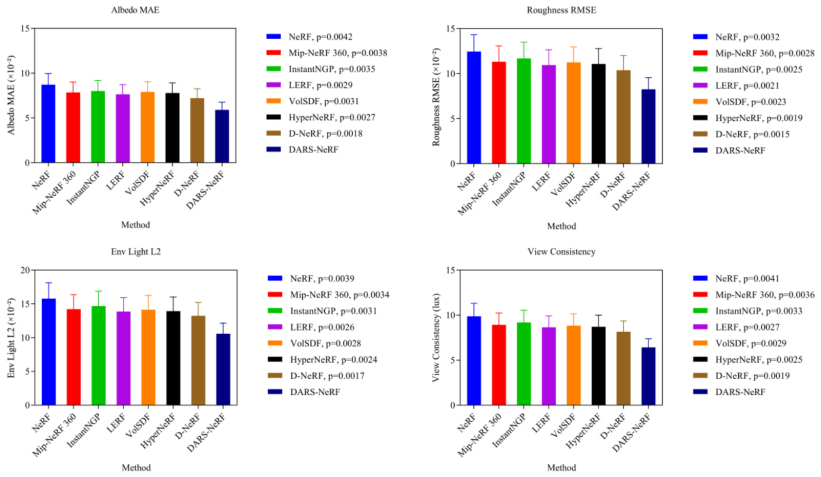

FIGURE 4. Statistical Significance of the Results

by coupled modeling. Furthermore, the DARS sampling strategy adds sampling points at occlusion boundaries and reflective areas, effectively capturing the light behavior in these key areas prone to viewpoint inconsistencies. This improves overall cross-view rendering consistency and provides a more stable visual experience for virtual roaming.

B. VISUALIZATION OF RECONSTRUCTION EFFECTS

The visual comparison of the reconstruction effects of different methods is shown in Figure 5.

DARS-NeRF's reconstruction results surpass comparable methods in visual fidelity and detail restoration, with output images closest to the original. This is primarily due to its dual-module collaborative design, which effectively addresses the fundamental shortcomings of traditional NeRF in art scenes. First, the distortion-aware sampling module dynamically weights semantic edges and gradient maps to increase sampling density in high-frequency regions such as highlights, transmission, and occlusion boundaries. This allows the network to fully learn the sudden radiometric changes at the interface between specular reflections and complex geometry, avoiding the blurring and artifacts often associated with undersampling in standard NeRF. Second, the semantically guided structure-material-lighting decoupling architecture explicitly separates PBR parameters from ambient lighting, ensuring that albedo is free of light contamination, roughness conforms to material physical priors, and metallic areas are precisely located. DARS-NeRF, through its integrated "sampling adaptation + physical decoupling + semantic constraints" mechanism, achieves high-fidelity coordinated reconstruction of geometry, materials, and lighting in art scenes without increasing model capacity. Currently, it is the only neural rendering framework to achieve "visually indistinguishable" quality under the complex lighting and material conditions found in real art galleries. The training loss is shown in Figure 6.

The training loss curves shown in Figure 6 demonstrate stable convergence of all loss terms in DARS-NeRF. The total loss Ltotal decreases rapidly within 5,000 steps and stabilizes after 10,000 steps, reflecting the model's efficient absorption of multimodal supervisory signals. The simultaneous decreases of Lrgb and Lssim indicate the coordinated optimization of pixel fidelity and structural consistency. Lmat converges quickly and early, thanks to semantically guided physical constraints that effectively prevent material parameter drift. Leikonal and Ldistortion maintain low values, indicating good control of geometric field smoothness and sampling weight stability. The negative correlation trend of Lpsnr confirms the consistency between the objective metrics and the loss function design. DARS-NeRF achieves convergence without inducing oscillations through a multi-loss synergy mechanism, demonstrating the superiority of its architectural design in optimizing dynamics, making it particularly suitable for high-dimensional, multi-constrained artistic scene reconstruction tasks.

C. ABLATION EXPERIMENT

It sets up an ablation experiment to analyze the impact of each component on performance. The results are shown in Table 6.

TABLE 6. Ablation Experiment Results

Model variants	PSNR (dB)	SSIM	LPIPS
Full DARS-NeRF	28.9	0.912	0.142
w/o DARS	26.7	0.873	0.189
w/o semantic guidance (Mask2Former)	27.1	0.881	0.176
w/o lighting decoupling (fixed lighting)	27.5	0.892	0.168
w/o material constraint loss	27.8	0.896	0.162
Standard NeRF	24.3	0.843	0.218

Note: w/o (without).

Ablation experiments show that each core module of DARS-NeRF significantly contributes to reconstruction performance. Removing the DARS sampling module reduces

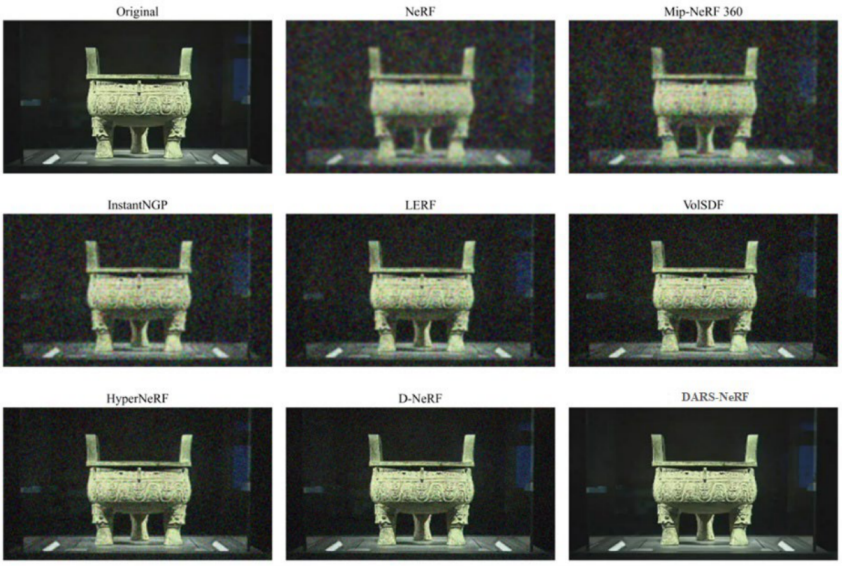

FIGURE 5. Visualization of Reconstruction Effect

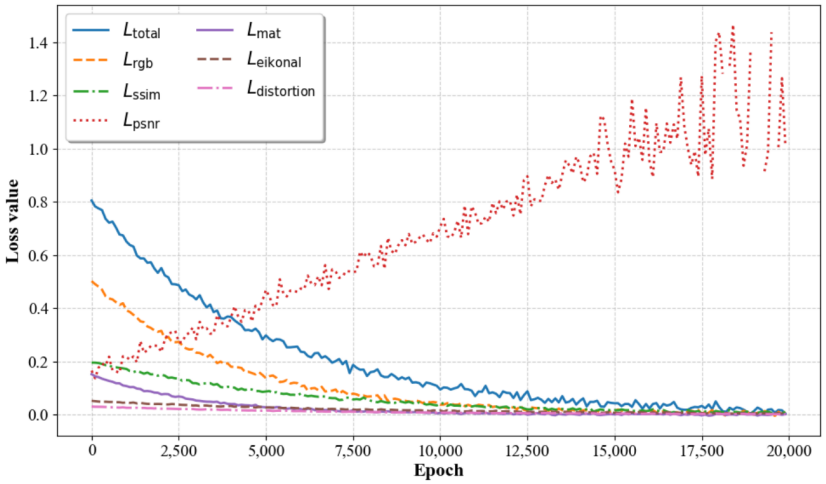

FIGURE 6. Training Loss Changes

PSNR by 2.2 dB and increases LPIPS by 33.1%, demonstrating that adaptive sampling is crucial for recovering detail in highlights and boundary regions. Disabling semantic guidance leads to material mismatch and a 1.8 dB decrease in PSNR, validating the effectiveness of semantic priors for constraining physical parameters. Fixed illumination head degrades cross-view consistency, resulting in a 0.020 decrease in SSIM, reflecting the necessity of illumination decoupling for structural fidelity. Removing the material loss term, while having a minor impact (PSNR down 1.1 dB), still results in color shifts in oil painting brushstrokes and metallic reflective areas, demonstrating the indispensability of physical regularization.

The sampling weight distribution of the artistic image is shown in Figure 7.

A comparative analysis of NeRF and DARS-NeRF sampling weight heatmaps reveals that standard NeRF employs a uniform stratified sampling strategy, resulting in a spatially homogenized weight distribution. This strategy lacks sampling enhancement in specular highlights, geometrically occluded boundaries, and translucent regions, leading to radiometric estimation bias in these high-frequency distortion regions due to sparse sampling. In contrast, DARS-NeRF generates a spatially adaptive sampling weight map through a weighted fusion of semantic edge maps, gradient magnitude maps, and metallic masks. This results in a significantly higher weight response in key areas such as reflective areas of picture frames and glass display case edges, driving increased sampling density, while actively reducing sampling in uniform walls or shadowed areas, achieving precise allocation of computing resources. This mechanism essentially samples the importance of the nonlinear integral term in the radiative transfer formula, focusing the limited sampling points on physical failure regions with the highest

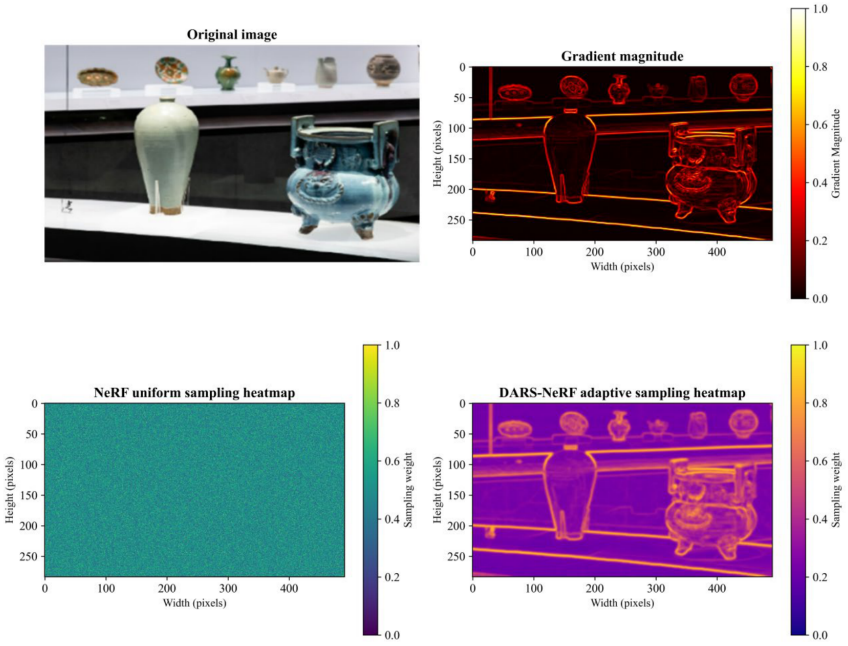

FIGURE 7. Sampling Weight Heat Map

information entropy and most drastic gradients. This significantly improves reconstruction stability and detail fidelity without increasing the total number of samples, demonstrating its irreplaceable superiority in artistic scenes with complex materials and non-uniform lighting conditions.

D. CROSS-VIEW ILLUMINATION ROBUSTNESS ANALYSIS

Cross-view lighting robustness analysis verifies the reconstruction method's ability to maintain lighting consistency across different observation angles. By measuring the brightness variations of the same surface at different viewpoints, it can quantitatively assess the effectiveness of the neural rendering system's decoupling of lighting and materials. Low brightness fluctuations indicate that the method successfully separates the scene's inherent reflective properties from the ambient lighting component, demonstrating physically realistic rendering capabilities and stable cross-view performance. The cross-view lighting robustness results are shown in Figure 8.

DARS-NeRF exhibits the best illumination consistency across different viewpoints, with minimal brightness fluctuation (92.9-116.5 lux). This demonstrates its successful decoupling of illumination from materials. Through explicit spherical harmonic illumination modeling and physically constrained loss, it avoids incorrectly encoding view-dependent illumination effects into material parameters. Traditional methods, such as NeRF, exhibit significant brightness fluctuations (85.1-134.9 lux), which stem from ambient light confounding caused by coupled modeling. DARS-NeRF's flat curve verifies the physical accuracy and viewpoint invariance of its illumination model, providing a stable visual experience for virtual tours and ensuring the realism and reliability of cultural heritage digitization.

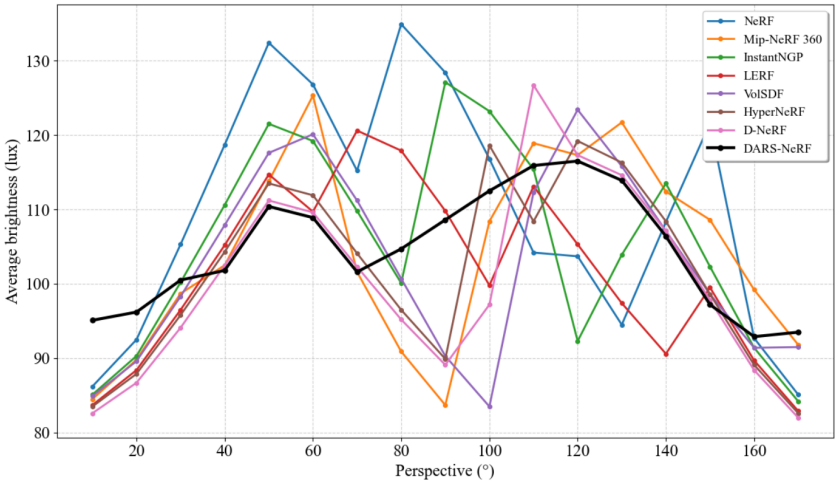

FIGURE 8. Cross-view Lighting Robustness

V. CONCLUSIONS

This paper proposes the DARS-NeRF framework, which integrates distortion-aware light sampling with structure-material-illumination decoupling modeling, to effectively address reconstruction distortion issues caused by complex material reflections and non-uniform illumination in virtual art gallery scenes, a methodology that is highly transferable to creating accurate digital twins of complex materials by isolating the effects of weave structure, fiber material, and viewing conditions. By dynamically adjusting the sampling den-sity distribution, it adaptively compensates for sampling bias in visually distorted areas caused by reflections, transmission, and occlusions. A multi-branch NeRF architecture is applied to explicitly decouple geometry, material, and illumination. A physically constrained loss and an alternating optimization strategy ensure training stability and physical consistency of the output. Experimental results demonstrate that DARS-NeRF achieves PSNR of 28.9 dB, SSIM of 0.912, and LPIPS of 0.142, respectively, significantly improving over the existing best baseline methods. It also demonstrates higher accuracy in Albedo and Roughness, and achieves optimal cross-view illumination consistency (luminance fluctuations between 92.9 and 116.5 lux). This study not only verifies the synergistic advantages of distortion-aware sampling and semantically guided decoupled modeling but also provides a new technical path for high-fidelity preservation, virtual exhibitions, and immersive communication of digital cultural heritage, expanding the application boundaries of neural radiation fields in complex art scenes.

AUTHOR CONTRIBUTIONS

Xiaopeng Pei designed, collected and analyzed the data, and drafted the manuscript. Xiaopeng Pei conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Xiaopeng Pei participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by,

1. The Research and Practice Project of Higher Education Reform in Henan Province in 2024, No.: 2024SJGLX0861
2. Research and Practice Project on Higher Education Teaching Reform in Henan Province in 2022, No. : 2021SJGLX787
3. School-based key project of Hebi Vocational and Technical College in 2024, No.:2024-KJZD-003
4. Special Research Project for Educational Strong Province in Henan Province (2026), No.: 2026JYQS062
5. Provincial Social Science Association Research Project for 2025, No.: SKL-2025-393

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] F. I. Apollonio, F. Fantini, S. Garagnani, and M. Gaiani, "A photogrammetry-based workflow for the accurate 3D construction and visualization of museums assets," Remote Sensing, vol. 13, no. 3, pp. 486, 2021, doi: 10.3390/rs13030486.

[2] Y. Wang, D. Sui, Y. Li, Y. Li, and M. Guo, "RAS-NeRF for novel view synthesis of complex reflective artifacts," npj Heritage Science, vol. 13, no. 1, pp. 161, 2025, doi: 10.1038/s40494-025-01727-6.

[3] E. M. Farella, L. Morelli, S. Rigon, E. Grilli, and F. Remondino, "Analysing key steps of the photogrammetric pipeline for Museum artefacts 3D digitisation," Sustainability, vol. 14, no. 9, pp. 5740, 2022, doi: 10.3390/su14095740.

[4] L. Gao, Y. Zhao, J. Han, and H. Liu, "Research on multi-view 3D reconstruction technology based on SFM," Sensors, vol. 22, no. 12, pp. 4366, 2022, doi: 10.3390/s22124366.

[5] T. Nakagawa, "Research on 3D Measurements used for Archaeological Materials in Japan," The Indonesian Journal of Social Studies, vol. 7, no. 2, pp. 292-299, 2024, doi: 10.26740/ijss.v7n2.p292-299.
[6] F. Remondino, A. Karami, Z. Yan, G. Mazzacca, S. Rigon, R. Qin, et al., "A critical analysis of NeRF-based 3D reconstruction," Remote Sensing, vol. 15, no. 14, pp. 3585, 2023, doi: 10.3390/rs15143585.
[7] V. Croce, D. Billi, G. Caroti, A. Piemonte, L. De Luca, P. Veron, et al., "Comparative assessment of neural radiance fields and photogrammetry in digital heritage: impact of varying image conditions on 3D reconstruction," Remote Sensing, vol. 16, no. 2, pp. 301, 2024, doi: 10.3390/rs16020301.
[8] C. Shi, H. Deng, S. Xiang, and J. Wu, "3D reconstruction based on fuzzy optimization of neural radiation field structure," Chinese Journal of Liquid Crystals and Displays, vol. 39, no. 11, pp. 1483, 2024, doi: 10.37188/CJLCD.2024-0155.
[9] T. Chen, Q. Yang, and Y. Chen, "A review of neural radiation field technology and applications," Journal of Computer-Aided Design & Graphics, vol. 37, no. 1, pp. 51-74, 2025, doi: 10.3724/SP.J.1089.2023-00759.
[10] C. Sun, J. Qiu, L. Wu, and C. Liu, "Dynamic human neural radiation field reconstruction based on monocular vision," Acta Optica Sinica, vol. 44, no. 19, Art. no. 1915001, 2024, doi: 10.3788/AOS240809.
[11] J. Li, L. Cheng, J. He, and Z. Wang, "Research status and prospects of neural radiation field," Journal of Computer-Aided Design & Computer Graphics, vol. 36, no. 7, pp. 995-1013, 2024, doi: 10.3724/SP.J.1089.2024.2023-00376.
[12] R. Wei, H. Pei, D. Wu, C. Zeng, X. Ai, and H. Duan, "A Semantically Aware Multi-View 3D Reconstruction Method for Urban Applications," Applied Sciences-Basel, vol. 14, no. 5, pp. 2218, 2024, doi: 10.3390/app14052218.
[13] Z. Xie, H. Liu, Y. He, Y. Shi, P. Yu, J. Ai, et al., "Cross modal networks for point cloud semantic segmentation of Chinese ancient buildings," npj Heritage Science, vol. 13, no. 1, pp. 131, 2025, doi: 10.1038/s40494-025-01701-2.
[14] R. Li, P. Dai, G. Liu, S. Zhang, B. Zeng, and S. Liu, "PBR-GAN: Imitating physically-based rendering with generative adversarial networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 3, pp. 1827-1840, 2023, doi: 10.1109/TCSVT.2023.3298929.
[15] M. Rossoni, M. Pozzi, G. Colombo, M. Gribaudo, and P. Piazzolla, "Physically based rendering of animated point clouds for extended reality," Journal of Computing and Information Science in Engineering, vol. 24, no. 5, Art. no. 054501, 2024, doi: 10.1115/1.4063559.
[16] Y. Zhang, Y. Liu, Z. Xie, L. Yang, Z. Liu, M. Yang, et al., "Dreammat: High-quality pbr material generation with geometry-and light-aware diffusion models," ACM Transactions on Graphics (TOG), vol. 43, no. 4, pp. 1-18, 2024, doi: 10.1145/3658170.
[17] B. Cai, Y. Li, Y. Liang, R. Jia, B. Zhao, M. Gong, et al., "3D Scene Creation and Rendering via Rough Meshes: A Lighting Transfer Avenue," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 9, pp. 6292-6305, 2024, doi: 10.1109/TPAMI.2024.3381982.
[18] Y. Peng, Y. Gao, T. Du, Y. Sang, and L. Zi, "Single image super-resolution reconstruction method based on generative adversarial network," Journal of Frontiers of Computer Science and Technology, vol. 14, no. 9, pp. 1612-1620, 2020, doi: 10.1038/s41598-025-28677-0.
[19] Y. Xin, F. Zhu, P. Shi, X. Yang, and R. Zhou, "Super-Resolution Reconstruction Algorithm of Images Based on Improved Enhanced Super-Resolution Generative Adversarial Network," Laser & Optoelectronics Progress, vol. 59, no. 4, Art. no. 0420002, 2022, doi: 10.3788/lop202259.0420002.