# AdaGA-NeRF for Sparse-View Synthesis

Fan Wan‡, Zhihan Jiang‡, Yuchen Li‡, Xingyu Miao, Xueqi Qiu, Tianyu Zhang, Yang Long ✉, *Senior Member IEEE*

***Abstract*—Neural Radiance Fields (NeRF) have transformed novel view synthesis by learning scene geometry and appearance from multi-view images. However, their reliance on densely sampled viewpoints makes them vulnerable under sparse-view conditions commonly found in real-world scenarios. In this work, we propose Adaptive Geometry-Aware NeRF (AdaGA-NeRF), a fully self-contained framework that overcomes these limitations without relying on external priors. First, we introduce an adaptive scene representation that leverages entropy-driven uncertainty to dynamically modulate depth constraints and focus supervision on ambiguous regions, addressing inconsistent density predictions in undersampled areas. We further develop a depth-consistent regularization, formulated via a spatially weighted KL-divergence, which enforces both local detail preservation and global geometric coherence to mitigate depth discontinuities. To address the low-frequency bias of standard MLP-based architectures, our Hierarchical Detail Refinement (HDR) module recovers high-frequency local details through nonlinear transformations and differential optimization, while our Structural Consistency Reinforcement (SCR) module adaptively recalibrates features to maintain global structural alignment, both motivated by the need to recover fine textures and avoid smoothing artifacts. Extensive experimental results demonstrate that AdaGA-NeRF achieves superior performance and robustness in sparse-view scenarios, significantly improving reconstruction quality and generalisation capabilities compared to state-of-the-art approaches.**

***Index Terms*—Neural Radiance Fields (NeRF), Novel View Synthesis, Sparse-View Synthesis, 3D Scene Reconstruction**

## I. Introduction

Recent advances in Neural Radiance Fields (NeRF) [1] have transformed novel view synthesis by leveraging deep neural networks to implicitly encode both geometric and photometric scene characteristics. This implicit representation enables photorealistic rendering of complex scenes from unseen viewpoints, finding applications in virtual reality (VR) [2], augmented reality (AR) [3], robotic navigation [4], [5], digital content creation [6], and autonomous driving [7]. However, NeRF's remarkable performance is largely predicated on densely sampled input images. In practical settings where only a limited number of views are available, a situation known as *sparse-view problem*, NeRF struggles with depth ambiguities, leading to geometric inconsistencies and visual artifacts such as blurred textures, floating structures, and overall structural distortions.

Fan Wan is with the Central Research Institute, Tongfang Knowledge Network Digital Technology Co., Ltd., China National Nuclear Corporation, Beijing, China. Yuchen Li, Xingyu Miao, Xueqi Qiu, Tianyu Zhang, Yang Long are with the Department of Computer Science, Durham University. (E-mail:fan.wan.uk@gmail.com; {yuchen.li; xingyu.miao; xueqi.qiu; tianyu.zhang; yang.long}@durham.ac.uk); Zhihan Jiang is with the Faculty of Arts and Social Sciences, National University of Singapore, Singapore (E-mail: e1373076@u.nus.edu)

‡ Equal contribution.

✉ Corresponding author: Yang Long(yang.long@durham.ac.uk).

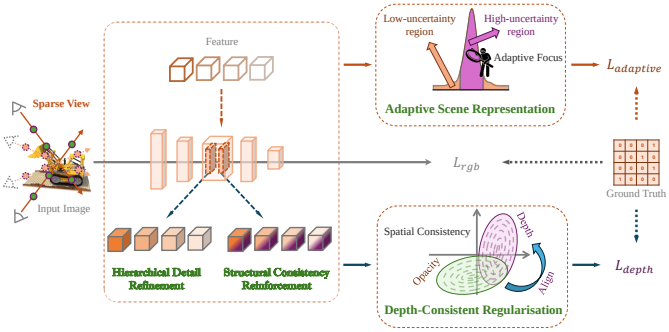


Fig. 1. Overview of the AdaGA-NeRF framework. Our approach integrates adaptive scene representation, depth-consistent regularization, hierarchical detail refinement, and structural consistency reinforcement to improve sparse-view reconstruction without external priors.

Prior approaches addressing sparse-view challenges often incorporate external priors, such as explicit depth cues or pre-trained visual knowledge, to enhance reconstruction quality. A widely adopted solution integrates depth priors to explicitly guide geometric understanding [8], [9], typically relying on preliminary procedures such as monocular depth estimation, depth map refinement, and geometric verification to provide additional structural cues. Although depth priors significantly improve the ability of NeRF to resolve depth ambiguities, they introduce computational overhead due to pre-processing requirements and may propagate estimation inaccuracies, particularly in scenes with occlusions, reflective surfaces, or complex geometries. Another approach employs pre-trained neural networks for feature transfer [10], [11], leveraging external visual knowledge to enhance NeRF's generalization. While effective, this strategy significantly increases computational complexity and introduces dataset-specific biases, limiting its applicability in diverse, resource-constrained environments. Recent works [12]–[14] have also explored explicit geometric regularizers, uncertainty-driven methods, and internal priors. However, these solutions either lack sufficient global–local balancing or still depend on external inputs, further motivating the need for a fully self-contained, geometry-adaptive solution.

To address these limitations, we propose the *Adaptive Geometry-Aware Neural Radiance Field* (AdaGA-NeRF), a fully self-contained framework that robustly handles sparse-view reconstruction without external priors. Unlike conventional NeRF methods that rely primarily on color-based supervision, AdaGA-NeRF introduces explicit geometric constraints derived directly from scene uncertainty distributions, allowing for adaptive supervision of ambiguous regions to ensure accurate scene reconstruction.

In environments characterized by simple geometry and uniform lighting, critical geometric information is primarily

concentrated near object surfaces. To capture this effectively, we introduce an entropy-based regularization approach that uses ray entropy to measure uncertainty in the opacity distribution along rays. Minimizing ray entropy encourages opacity distributions to concentrate near object boundaries, thereby enhancing geometric sharpness and accuracy. However, this strategy alone is insufficient for complex scenes where intricate geometries, varying illumination, reflections, and diverse materials can produce scattered and noisy geometric signals.

To address this challenge, we introduce a spatially weighted depth consistency constraint based on the Kullback–Leibler (KL) divergence, ensuring both local detail preservation and global geometric coherence. This additional regularization distributes supervision more evenly across the scene, mitigating depth discontinuities and reducing floating artifacts. To dynamically balance these constraints, we incorporate an adaptive weighting strategy based on estimated depth uncertainty, enabling robust performance across diverse environments.

Despite these adaptive geometric strategies, standard multilayer perceptron (MLP)-based architectures in NeRF tend to favor smooth, low-frequency representations, often resulting in oversmoothed textures and blurred edges under sparse-view conditions [15]. Furthermore, these architectures lack explicit mechanisms to enforce global structural consistency across synthesized views, which exacerbates floating artifacts and misalignment issues [16], [17]. To address these inherent limitations, AdaGA-NeRF introduces two architectural innovations. The *Hierarchical Detail Refinement* (HDR) module enhances high-frequency local details through progressive nonlinear feature transformations and differential optimization, achieving finer detail recovery distinct from conventional feature refinement methods such as [18]–[20]. Meanwhile, the *Structural Consistency Reinforcement* (SCR) module leverages adaptive channel-wise feature recalibration to strengthen global structural alignment across synthesized views. The synergy between the HDR and SCR modules enables AdaGA-NeRF to simultaneously achieve high-fidelity detail recovery and structural consistency under sparse-view conditions.

By integrating these adaptive geometric constraints with complementary architectural enhancements, AdaGA-NeRF achieves robust and accurate reconstruction under sparse-view conditions while eliminating the need for external priors. Figure 1 provides an overview of the proposed framework. Our primary contributions are summarized as follows:

- **Adaptive Scene Representation:** An uncertainty-driven geometric supervision mechanism that dynamically modulates constraints based on ray entropy and depth uncertainty.
- **Depth-Consistent Regularization:** A spatially weighted KL-divergence regularization that enforces both local detail preservation and global geometric coherence.
- **Hierarchical Detail Refinement with Structural Consistency Reinforcement:** A unified architectural innovation that progressively recovers high-frequency local details through nonlinear feature transformations while simultaneously enhancing global structural alignment via adaptive channel-wise recalibration.

## II. Related Work

**Novel View Synthesis and NeRF.** Novel view synthesis is a long-standing problem in computer vision [21]–[27] that aims to render unseen viewpoints of a scene from a limited set of images. Neural Radiance Fields (NeRF) [1], [28] achieved breakthrough results in this task by representing a scene as a continuous volumetric function learned from many posed images. However, NeRF's performance degrades severely when the input views are sparse. Capturing dozens of images per scene is often impractical, making sparse-view (few-shot) novel view synthesis a core challenge. Consequently, recent research has explored various priors and regularizations to make NeRFs work with far fewer images.

**Depth-Prior Based Methods.** A prominent direction for improving NeRF under sparse supervision is the incorporation of geometric depth priors. Depth-supervised NeRF (DS-NeRF) [9] first demonstrated that even very sparse depth cues from structure-from-motion (SfM) can significantly improve novel view synthesis with few input images. By supervising ray termination depth using sparse SfM point clouds, DS-NeRF reduces floaters and accelerates convergence. Building on this, Roessle et al. [29] employ dense depth maps obtained via depth completion to further constrain scene geometry, achieving high-fidelity reconstruction with fewer views.

To relax reliance on accurate metric depth, SparseNeRF [8] introduces relative depth rankings distilled from consumer-grade RGB-D sensors or monocular depth prediction networks. These weak priors enforce multi-view consistency without needing precise ground-truth depth.

Similarly, the RGB-D-guided approach in Neural Radiance Fields From Sparse RGB-D Images [30] utilizes mesh-based geometry derived from depth data to pre-train NeRF via pseudo-ground-truth rendering, offering strong supervision while retaining flexibility.

Despite their effectiveness, depth-prior methods heavily depend on external sensors or depth-estimation models, which may not always be reliable or available. They also generally require per-scene depth optimization, limiting generalization to novel environments lacking depth annotations.

**Pre-Trained Feature and Distillation Methods.** To address data scarcity and enable faster generalization, another class of work leverages pre-trained models or learns scene priors from large-scale datasets. PixelNeRF [31] and MVSNeRF [32] are early examples, training NeRF-based models across multiple scenes so that at inference they can synthesize novel views from as few as one or three images. PixelNeRF conditions NeRF on image-level CNN features, learning cross-scene priors, while MVSNeRF introduces a plane-sweep cost volume to infer geometry with limited inputs.

MorphNeRF [37] extends these ideas by integrating CLIP-based semantic guidance into a morphing field, enabling text-driven or exemplar-guided geometry-aware editing without per-scene retraining. More recent efforts incorporate foundation models: CLIP-NeRF [11] uses a CLIP-based semantic loss for NeRF editing, and DINO-NeRF [33] distills self-supervised ViT features for regularization. NeRF Signature [38] also harnesses codebook-driven priors, embedding robust watermark signals without compromising visual fidelity.

TABLE I
COMPARISON OF REPRESENTATIVE SPARSE-VIEW NERF METHODS. "EXTERNAL INPUT/MODEL" DENOTES ANY ADDITIONAL INPUTS (BEYOND RGB IMAGES) OR PRE-TRAINED MODELS REQUIRED. "EFFICIENCY / COMPLEXITY" REFLECTS TRAINING/INFERENCE OVERHEAD AND DEPLOYMENT PRACTICALITY.

| Method (Year) | External Input / Model | Key Characteristics (Remarks) | Efficiency / Complexity |
|---|---|---|---|
| NeRF [1] (2020) | ✗ | Baseline method; requires dense view coverage (50+ views). | Moderate training cost; no external dependency; not optimized for sparse input. |
| DS-NeRF [9] (2022) | Depth (SfM) | Uses sparse SfM depth points for improved supervision. | Low overhead; relies on COLMAP; faster convergence than NeRF. |
| DenseNeRF [29] (2022) | Dense depth | High-quality reconstruction using completed depth maps. | High preprocessing cost; slower training; complex deployment. |
| SparseNeRF [8] (2023) | Coarse depth | Depth ranking from monocular priors; robust to noise. | Lightweight; needs external depth; easy to integrate. |
| PixelNeRF [31] (2021) | ✗ | Pre-trained CNN; generalizes across scenes. | Fast inference; high training cost; large model size. |
| MVSNeRF [32] (2021) | ✗ | Plane-sweep volumes; leverages multi-view stereo. | heavy architecture; high memory demand. |
| CLIP-NeRF [11] (2022) | CLIP | Text-guided NeRF editing via CLIP embeddings. | Very high training cost; not scalable; multiple large models. |
| DINO-NeRF [33] (2022) | DINO | Semantic consistency via ViT features. | High overhead during training; depends on ViT. |
| CustomNeRF [34] (2024) | Diffusion model | Scene editing guided by diffusion priors. | Extremely complex; slow optimization; GPU-intensive. |
| RegNeRF [16] (2022) | ✗ | Uses depth continuity and patch realism constraints. | Medium overhead; needs normalizing flow model. |
| InfoNeRF [35] (2022) | ✗ | Ray entropy regularization to reduce floaters. | Low overhead; self-contained; easy to scale. |
| DietNeRF [10] (2021) | ✗ | DietNeRF uses self-supervised feature alignment for cross-view semantic consistency. CLIP may be optionally used but is not essential. | Moderate training cost; depends on frozen CLIP features. |
| FreeNeRF [36] (2023) | ✗ | Frequency regularization prevents overfitting. | Minimal cost; very efficient; no extra components. |
| AdaGA-NeRF (Ours) | ✗ | Adaptive geometric constraints (entropy & depth) without priors. | Lightweight; no external data; moderate training overhead only. |

While these approaches facilitate rapid inference and leverage powerful semantic priors, they often involve large architectures and higher training overhead, potentially underperforming carefully optimized per-scene NeRF methods in fine-grained detail and accuracy.

**Information-Theoretic and Intrinsic Regularization.** Complementary to external priors, several methods introduce intrinsic constraints into NeRF optimization to improve few-shot performance. RegNeRF [16] adds depth and appearance regularizers, encouraging geometry continuity through flow-based warping. InfoNeRF [35] imposes entropy loss over volumetric density to reduce floaters and artifacts. SID-NeRF [28] further leverages information-theoretic cues by jointly minimizing ray-entropy and a depth-difference–weight KL divergence, while a selector–residual mechanism mitigates noisy supervision. DietNeRF [10] proposes self-supervised feature alignment for semantic consistency across views. ATM-NeRF [39] enhances intrinsic regularization with self-generated geometric supervision, offering geometry-aware constraints in resource-limited setups.

These intrinsic methods are lightweight and broadly applicable; however, their generic priors (e.g., smoothness, entropy) can struggle with scenes exhibiting complex geometry or intricate details, limiting their efficacy in challenging sparse-view scenarios.

**Geometry-Aware Internal Regularization.** Another approach enforces internal geometric consistency without external data. Multi-view feature warping [40] helps prevent view-specific overfitting, while CBARF [41] uses cascaded bundle adjustment and spatial pose correction to align geometry and appearance. OM-NeRF [42] integrates 3D-aware attention and SMPL-based occlusion modeling for improved human-centric NeRF under severe self-occlusions. FreeNeRF [36] applies frequency regularization to avoid overfitting sparse inputs.

Our proposed *AdaGA-NeRF (Adaptive Geometry-Aware NeRF)* follows this direction by combining internal geometry-attention and adaptive regularizers to guide the radiance field towards underlying structural cues. Like other self-contained methods, it does not rely on external data. By adaptively sampling spatial frequencies and enforcing depth consistency, AdaGA-NeRF reconstructs more accurate surfaces from limited images while maintaining efficient inference.

**Discussion and Summary.** Table I highlights representative sparse-view NeRF methods. Although external priors can improve reconstruction, they increase complexity and can limit real-world applicability. Intrinsic regularization offers flexible,

lightweight solutions but may struggle with extremely sparse or complex scenes.

Recent trends combine geometric constraints, learned priors, and adaptive regularization to achieve superior quality. AdaGA-NeRF fits into this movement by integrating adaptive geometric constraints without relying on external data, effectively narrowing the performance gap in sparse-view scenarios.

## III. METHODOLOGY

### A. NeRF Framework and Sparse-View Challenges

Neural Radiance Fields (NeRF) represent a scene as a continuous volumetric function parameterized by a neural network. Given a 3D coordinate $\mathbf{x} = (x, y, z)$ and a viewing direction $(\theta, \phi)$, NeRF predicts the RGB color $\mathbf{c} = (r, g, b)$ and density $\sigma$. These predictions are integrated along a ray using classical volume rendering:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \left(1 - \exp(-\sigma_i \delta_i)\right) \mathbf{c}_i, \quad T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right), \tag{1}$$

where $T_i$ denotes the accumulated transmittance up to sample $i$, and $\delta_i$ is the interval between samples. NeRF's success in synthesizing photorealistic images relies on sufficient multi-view samples that provide strong geometric constraints for accurate density prediction and surface reconstruction.

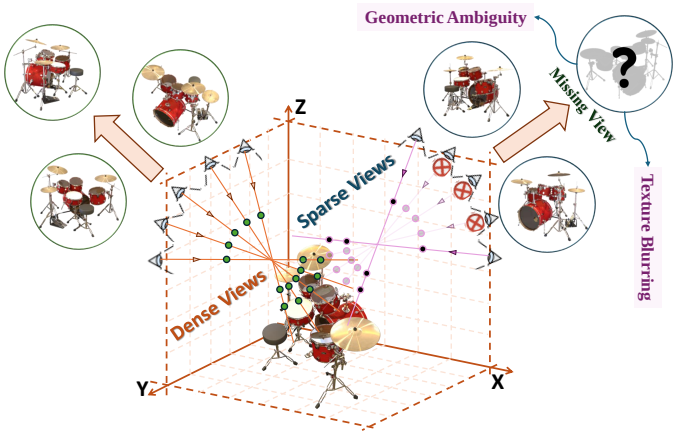


Fig. 2. Motivation behind AdaGA-NeRF. Sparse-view NeRF suffers from incomplete geometric supervision and missing viewpoints, leading to geometric ambiguity and texture blurring. Compared to dense-view setups, sparse inputs introduce severe depth uncertainty and incomplete scene coverage. This motivates our geometry-aware, uncertainty-driven framework for robust novel view synthesis under sparse conditions.

In sparse-view settings, however, limited viewpoints lead to under-constrained density estimation, resulting in two primary challenges (Fig. 2): (1) *geometric ambiguity*—manifested as floating artifacts due to uncertain density estimates, and (2) *texture blurring*—caused by inadequate inter-ray constraints that hinder fine-detail recovery. While methods like InfoNeRF [35] mitigate these issues via per-ray entropy minimization, such local regularization lacks the spatial coherence needed to enforce global consistency, often yielding depth discontinuities and isolated artifacts in complex scenes.

These challenges motivate our development of Adaptive Scene Representation and Depth-Consistent Regularisation, which explicitly address local ambiguity and enhance global coherence, respectively, for robust novel view synthesis under sparse conditions.

### B. Adaptive Scene Representation

To tackle the challenges of NeRF reconstruction under sparse-view conditions, we propose an adaptive scene representation strategy that dynamically adjusts the sampling and optimization process based on geometric uncertainty. Unlike traditional NeRF methods that apply uniform sampling and supervision across all rays, our approach uses an uncertainty-aware weighting scheme to focus on regions with higher ambiguity, improving generalization and efficiency. We define the adaptive loss function as:

$$\mathcal{L}_{\text{adaptive}} = \min\left(\sum_{\mathbf{r}\in\mathcal{R}} \frac{H(\mathbf{r})}{1 + e^{-\beta(D(\mathbf{r}) - \mu_D)}}\right), \tag{2}$$

where $\mathcal{R}$ is the set of rays sampled from input viewpoints during training, $H(\mathbf{r})$ denotes the entropy-based uncertainty along ray $\mathbf{r}$, $D(\mathbf{r})$ is the estimated depth along the ray, $\mu_D$ is the global mean depth across all rays, and $\beta$ is a hyperparameter controlling the weighting function's sensitivity.

*1) Entropy-Based Uncertainty:* The entropy term $H(\mathbf{r})$ measures uncertainty in the density distribution along a ray, acting as an indicator of geometric ambiguity. Inspired by InfoNeRF [35], it is computed as:

$$H(\mathbf{r}) = -\sum_{i=1}^{N} p_i \log p_i,$$
$$\text{where} \quad p_i = \frac{\alpha_i}{\sum_{j=1}^{N} \alpha_j}, \quad \alpha_i = 1 - \exp(-\sigma_i \delta_i), \tag{3}$$

where $\sigma_i$ and $\delta_i$ representing the predicted density and sampling interval at the $i$-th sample along the ray, respectively, and $N$ being the number of samples. Here, $\alpha_i$ is the opacity at each sample, normalized into a probability distribution $p_i$, where high entropy reflects a diffuse density distribution and greater geometric uncertainty.

*2) Depth Estimation and Global Context:* The depth $D(\mathbf{r})$ along a ray is computed as the expected depth, weighted by opacity:

$$D(\mathbf{r}) = \sum_{i=1}^{N} w_i z_i, \quad \text{where} \quad w_i = \frac{\alpha_i}{\sum_{j=1}^{N} \alpha_j}, \tag{4}$$

with $z_i$ as the depth of the $i$-th sample. The global mean depth $\mu_D$ is dynamically calculated across all rays in a training batch: $\mu_D = \frac{1}{|\mathcal{R}|} \sum_{\mathbf{r}\in\mathcal{R}} D(\mathbf{r})$, ensuring adaptability to different scene scales without manual tuning.

*3) Adaptive Weighting Mechanism:* The weighting function $\frac{1}{1+e^{-\beta(D(\mathbf{r})-\mu_D)}}$ adjusts the influence of entropy regularization based on depth deviations from the mean. When $D(\mathbf{r}) \approx \mu_D$, the weight is approximately 0.5, providing balanced supervision; for rays with $D(\mathbf{r})$ significantly above or below $\mu_D$,

the weight shifts to emphasize regions with greater depth uncertainty, such as complex or sparse areas. This adaptive mechanism allows the model to dynamically prioritize regions with higher geometric uncertainty, enhancing the overall reconstruction quality.

By prioritizing supervision in geometrically ambiguous regions, our adaptive approach enhances density estimation and spatial coherence, reducing artifacts like depth discontinuities, which is crucial for real-world applications with limited input data where accurate geometry is essential.

### C. Depth-Consistent Regularization

Entropy-based regularization effectively concentrates opacity along individual rays but lacks inherent mechanisms to enforce spatial coherence across neighboring rays. This limitation often manifests as local inconsistencies in depth estimation, leading to fragmented or floating artifacts—particularly in complex scenes under sparse-view conditions. To mitigate this, we propose a depth-consistent regularization approach that aligns the predicted depth distribution with a spatially aware constraint, fostering structural continuity throughout the scene. We formulate our depth-consistent regularization as follows:

$$\mathcal{L}_{\text{depth}} = \sum_{\mathbf{r}\in\mathcal{R}} D_{\text{KL}}(w_r \,||\, d_r)\, e^{-\lambda|D(\mathbf{r})-\mu_D|}, \tag{5}$$

where $\mathcal{R}$ is the set of rays sampled from input viewpoints during training, $D_{\text{KL}}(w_r \,||\, d_r)$ denotes the Kullback-Leibler (KL) divergence between the predicted weight distribution $w_r$ and the depth distribution $d_r$, $D(\mathbf{r})$ is the expected depth along ray $\mathbf{r}$, computed as $D(\mathbf{r}) = \sum_i w_{r_i} z_i$ with $z_i$ representing the depth of the $i$-th sample along the ray, $\mu_D$ is the global mean depth across all rays in the training batch, defined as $\mu_D = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{r}\in\mathcal{R}} D(\mathbf{r})$, and $\lambda$ is a hyperparameter that modulates the sensitivity of the depth-adaptive weighting.

The predicted weight distribution $w_r$ is derived from the normalized opacity values of sampled points along each ray:

$$w_{r_i} = \frac{\alpha_i}{\sum_j \alpha_j}, \quad \text{where} \quad \alpha_i = 1 - \exp(-\sigma_i \delta_i), \tag{6}$$

with $\sigma_i$ representing the predicted density at the $i$-th sample and $\delta_i$ the interval between consecutive samples.

The depth distribution $d_r$ is obtained by normalizing the depth values of the sampled points using a softmax function:

$$d_{r_i} = \frac{e^{-z_i}}{\sum_j e^{-z_j}}, \tag{7}$$

where $z_i$ is the depth of the $i$-th sample along the ray.

The KL divergence $D_{\text{KL}}(w_r \,||\, d_r)$ quantifies the mismatch between the predicted weight distribution $w_r$ and the depth distribution $d_r$. This term ensures that the density-derived opacity aligns with the geometric structure of the scene. The exponential weighting factor $e^{-\lambda|D(\mathbf{r})-\mu_D|}$ introduces a depth-adaptive mechanism that intensifies regularization for rays whose expected depth $D(\mathbf{r})$ deviates significantly from the global mean $\mu_D$.

This adaptive weighting prioritizes supervision in regions with high depth uncertainty, such as those susceptible to floating artifacts or depth discontinuities, enhancing the model's ability to resolve such ambiguities.

By integrating this spatially weighted KL divergence, our depth-consistent regularization enforces structural coherence, reducing isolated artifacts. Unlike traditional per-ray depth constraints, our method leverages KL divergence and depth-adaptive weighting to enforce cross-ray spatial consistency, significantly enhancing structural coherence in sparse-view scenarios. This improves NeRF's robustness in sparse-view settings, leading to sharper reconstructions and better generalization to unseen viewpoints.

**Algorithm 1** Training Process of AdaGA-NeRF

**Require:** Training dataset with color data $I$, sampled ray set $\mathcal{R}$, learning rate $\eta$, number of iterations $N$
**Ensure:** Optimized model parameters $\theta$

Initialize model parameters $\theta$
**for** $n = 1$ to $N$ **do** ▷ Training iterations
  Predict initial features $h$, color $\hat{I}$, and density $\sigma$ at sampled points along rays in $\mathcal{R}$ using volumetric rendering.
  Apply HDR module to enhance features at multiple MLP layers:
$$h_{\text{HDR}} = h + \tanh(h \cdot h_{\text{pre}}) + (h - h_{\text{pre}})^2$$
  Apply the SCR module for global feature recalibration:
$$h_c = W_1 h_{\text{HDR}} + b_1, \quad h_{\text{SCR}} = h_{\text{HDR}} + W_2 \cdot \text{ReLU}(h_c) + b_2$$
  Recompute refined color and density predictions ($\hat{I}_{refined}$, $\sigma_{refined}$) using enhanced features $h_{\text{SCR}}$.
  Compute color reconstruction loss with refined predictions:
$$\mathcal{L}_{rgb} = \|\hat{I}_{refined} - I\|^2$$
  Evaluate adaptive scene representation loss:
$$\mathcal{L}_{adaptive} = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{r}\in\mathcal{R}} \frac{H(\mathbf{r})}{1 + e^{-\beta(D(\mathbf{r})-\mu_D)}}$$
  Compute depth-consistent regularization loss:
$$\mathcal{L}_{depth} = \frac{1}{|\mathcal{R}|}\sum_{\mathbf{r}\in\mathcal{R}} D_{KL}(w_r||d_r) e^{-\lambda|D(\mathbf{r})-\mu_D|}$$
  Aggregate total loss:
$$\mathcal{L}_{total} = \mathcal{L}_{rgb} + \mathcal{L}_{adaptive} + \mathcal{L}_{depth}$$
  Update model parameters $\theta$:
$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{total}$$
**end for**
**return** trained model parameters $\theta$

### D. Hierarchical Detail Refinement

Reconstructing fine-grained local details, such as sharp edges, intricate textures, and subtle geometric variations, poses a significant challenge for NeRF in sparse-view settings. The limited availability of multi-view constraints often leads to insufficient guidance for capturing high-frequency components. While adaptive geometric supervision and depth-consistent regularization offer optimization-level improvements, the standard multilayer perceptron (MLP) architecture of NeRF inherently biases toward smooth, low-frequency representations. This results in blurred or oversmoothed reconstructions, partic-

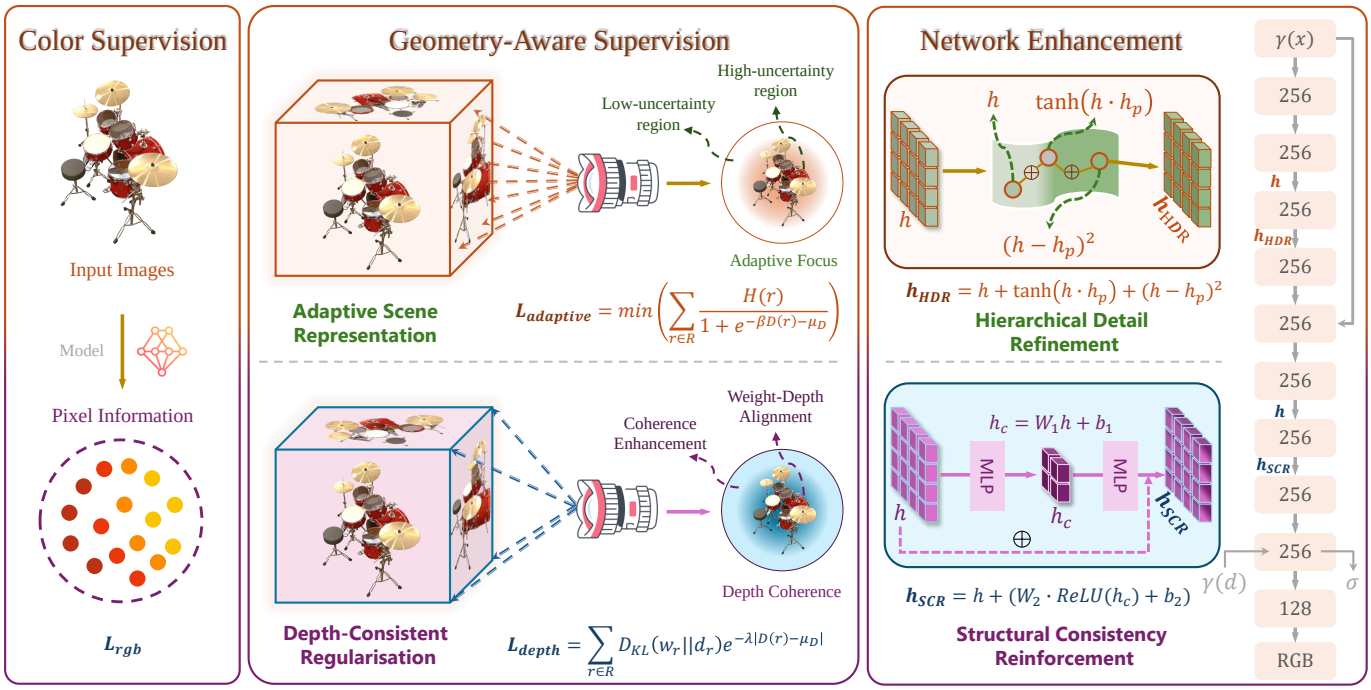


Fig. 3. Overview of the AdaGA-NeRF framework. The model combines color supervision, geometry-aware supervision, and network enhancements to address sparse-view challenges. Adaptive Scene Representation focuses constraints on uncertain regions via entropy-weighted depth, while Depth-Consistent Regularization aligns geometry across rays. Hierarchical Detail Refinement (HDR) enhances high-frequency details, and Structural Consistency Reinforcement (SCR) ensures global coherence, jointly improving reconstruction accuracy and structural consistency.

ularly in regions with complex details. To overcome this limitation, we propose a Hierarchical Detail Refinement (HDR) Module, a novel approach designed to explicitly enhance the network's capacity to represent high-frequency details across multiple scales.

The HDR module integrates two complementary feature refinement strategies: nonlinear feature enhancement and differential feature optimization. The nonlinear enhancement leverages a feature-wise transformation to amplify significant local variations, defined as:

$$\tanh(h \cdot h_{\text{pre}}), \tag{8}$$

where $h$ and $h_{\text{pre}}$ denote features from the current and preceding layers, respectively. The hyperbolic tangent (tanh) is chosen for its ability to model nonlinear interactions between features, selectively emphasizing high-frequency details, such as sharp gradient transitions or subtle illumination changes, which are typically underrepresented in standard MLPs. Compared to ReLU or GELU, tanh provides bounded and zero-centered outputs, which mitigate excessive noise amplification in sparse-view settings and preserve contrastive information from both positively and negatively correlated features. By amplifying these variations while suppressing redundant or noisy signals, this operation strengthens the sensitivity of the network to fine details.

In parallel, the differential optimization strategy targets rapid changes and high-frequency structures by computing a second-order difference between adjacent layers, expressed as:

$$(h - h_{\text{pre}})^2. \tag{9}$$

The squared difference term $(h - h_{\text{pre}})^2$ effectively acts as a nonlinear high-pass filter along the layer depth: the more abrupt the change between adjacent features, the larger this term becomes, thereby selectively enhancing sharp transitions in geometry or texture. By amplifying such inter-layer discrepancies, this term helps preserve fine-grained details that are otherwise prone to being smoothed out by standard MLPs.

The refined feature, $h_{\text{HDR}}$, combines the original feature with these two enhancements:

$$h_{\text{HDR}} = h + \tanh(h \cdot h_{\text{pre}}) + (h - h_{\text{pre}})^2. \tag{10}$$

This operation is applied hierarchically across the MLP layers, progressively refining local representations and improving the reconstruction of complex scene details. The hierarchical nature of the HDR module ensures that enhancements propagate across multiple scales, enhancing expressiveness in both fine and coarse regions of the scene.

### E. Structural Consistency Reinforcement

Ensuring global structural consistency is vital for accurate scene reconstruction in sparse-view conditions, where limited inputs often lead to depth discontinuities, misaligned surfaces, or floating artifacts. While the HDR Module improves fine-grained details, it does not inherently ensure coherent global structure, which is critical in sparse-view scenarios. To explicitly enhance global structural coherence, we introduce a Structural Consistency Reinforcement (SCR) Module, inspired by squeeze-and-excitation networks.

The proposed SCR Module consists of a two-step feature recalibration process. Firstly, input features $h$ undergo dimen-

sionality reduction, capturing the most relevant global patterns, described mathematically as:

$$h_c = W_1 h + b_1, \tag{11}$$

where $W_1 \in \mathbb{R}^{\frac{W}{k} \times W}$ and $b_1 \in \mathbb{R}^{\frac{W}{k}}$ are trainable parameters. This linear transformation efficiently recodes features, emphasizing critical global aspects of the scene.

Subsequently, the compressed features $h_c$ are activated through a nonlinear function and expanded back to the original dimensionality using another linear transformation:

$$h_{\text{SCR}} = h + W_2 \cdot \text{ReLU}(h_c) + b_2, \tag{12}$$

where $W_2 \in \mathbb{R}^{W \times \frac{W}{k}}$ and $b_2 \in \mathbb{R}^W$. This operation adaptively scales and activates globally significant features. A residual connection is incorporated to preserve local context and ensure stable gradient propagation during training. By adaptively reweighting channel activations, SCR amplifies salient global structures while retaining the fine-grained details preserved by HDR. Taken together, SCR derives channel-wise gating signals from aggregated cross-view features, uniformly recalibrates local representations, and guides the network toward coherent scene-level geometry.

### F. Integrated Training Framework

Our framework unifies adaptive geometric supervision and feature enhancements into a concise optimization scheme, enhancing NeRF's performance in sparse-view scenarios. The total loss integrates photometric reconstruction, adaptive geometric regularization, and depth consistency:

$$\mathcal{L}_{total} = L_{rgb} + \mathcal{L}_{adaptive} + \mathcal{L}_{depth}, \tag{13}$$

where $L_{rgb}$ is the photometric loss, $\mathcal{L}_{adaptive}$ targets uncertain regions using entropy weighting, and $\mathcal{L}_{depth}$ ensures depth coherence via spatially weighted KL-divergence. The HDR and SCR modules are embedded in the network and optimized jointly via gradient descent, enabling AdaGA-NeRF to capture fine details and maintain global consistency, outperforming prior methods. Overall, the pipeline of HDR, SCR, $\mathcal{L}_{adaptive}$, and $\mathcal{L}_{depth}$ forms a point-to-surface synergy chain. The HDR Module progressively amplifies and retains local high-frequency details within the network, while the SCR Module recalibrates these details in the channel dimension with global context, enforcing cross-view consistency. During training, the adaptive loss $\mathcal{L}_{adaptive}$ automatically concentrates supervision on rays with the highest geometric uncertainty whereas the subsequent depth-alignment loss $\mathcal{L}_{depth}$ uses a spatially consistent KL constraint to align the depth distributions of rays. In tandem, the front-end pair (HDR/SCR) augments the network's capacity for high-frequency representation and global structural coherence, and the back-end pair ($\mathcal{L}_{adaptive}$/$\mathcal{L}_{depth}$) dynamically steers gradients toward regions most prone to artifacts, ultimately delivering both sharp details and coherent global geometry under sparse-view conditions. Figure 3 outlines the framework, and Algorithm 1 details the training process.

## IV. EXPERIMENTS

### A. Setup

**Baselines and Comparisons.** We primarily benchmark AdaGA-NeRF against InfoNeRF [35], due to its entropy-based geometric regularisation explicitly designed for sparse-view scenarios. Additionally, we include comparisons with several mainstream state-of-the-art (SOTA) NeRF-based methods to comprehensively assess our method's performance.

**Datasets.** To evaluate the generalisation capability and effectiveness of AdaGA-NeRF under different conditions, we utilise three well-established NeRF benchmarks: the NeRF LLFF dataset [43], which features real-world forward-facing scenes with sparse viewpoints, presenting challenges like complex geometry and varied lighting conditions; the NeRF Synthetic dataset [1], composed of controlled computer-generated scenes ideal for evaluating precise geometric reconstruction fidelity; and the NeRF Real 360 dataset [1], containing complex real-world images from full 360-degree perspectives, which rigorously tests the model's capability to handle extensive viewpoint variations.

**Evaluation Metrics.** We quantitatively measure the novel view synthesis quality using three widely recognised metrics: Peak Signal-to-Noise Ratio (PSNR), assessing pixel-level accuracy through mean squared errors; Structural Similarity Index Measure (SSIM), evaluating spatial coherence and structural fidelity; and Learned Perceptual Image Patch Similarity (LPIPS), a perceptual metric based on deep network features reflecting human visual similarity. Higher PSNR and SSIM scores indicate better reconstruction accuracy, whereas lower LPIPS values signify improved perceptual realism.

**Implementation Details.** AdaGA-NeRF is implemented in PyTorch on an NVIDIA RTX 3090 GPU, using the Adam optimizer with an initial learning rate of $5 \times 10^{-4}$ and exponential decay. For the NeRF Synthetic and LLFF datasets, we use 4 views, and for the NeRF Real 360 dataset, we evaluate with both 4 and 8 views to ensure consistent evaluation across diverse viewpoint constraints. The hyperparameters are set to $\beta = 1$ and $\lambda = 0.1$, based on the empirical validation results.

### B. Main Results

**1) NeRF Synthetic Dataset Evaluation.** We first evaluate AdaGA-NeRF on the NeRF synthetic dataset [1], a widely used benchmark with precisely defined geometry and lighting. This dataset is particularly suited for assessing sparse-view reconstruction quality due to its controlled conditions. As shown in Table II, AdaGA-NeRF achieves state-of-the-art performance with a PSNR of $19.27 \pm 0.35$ dB and an SSIM of $0.813 \pm 0.006$, outperforming competing methods such as InfoNeRF ($18.65 \pm 0.18$ dB, $0.811 \pm 0.008$), DNGaussian ($18.96 \pm 1.23$ dB, $0.798 \pm 0.014$), and FreeNeRF ($19.03 \pm 1.52$ dB, $0.806 \pm 0.012$).

**Quantitative Analysis.** The high PSNR and SSIM values of AdaGA-NeRF indicate excellent pixel-level accuracy and structural fidelity, with the low standard deviation in PSNR ($\pm 0.35$) demonstrating its robustness across diverse scenes. Although both DNGaussian and FreeNeRF achieve slightly better LPIPS scores (with DNGaussian at $0.169 \pm 0.015$ and

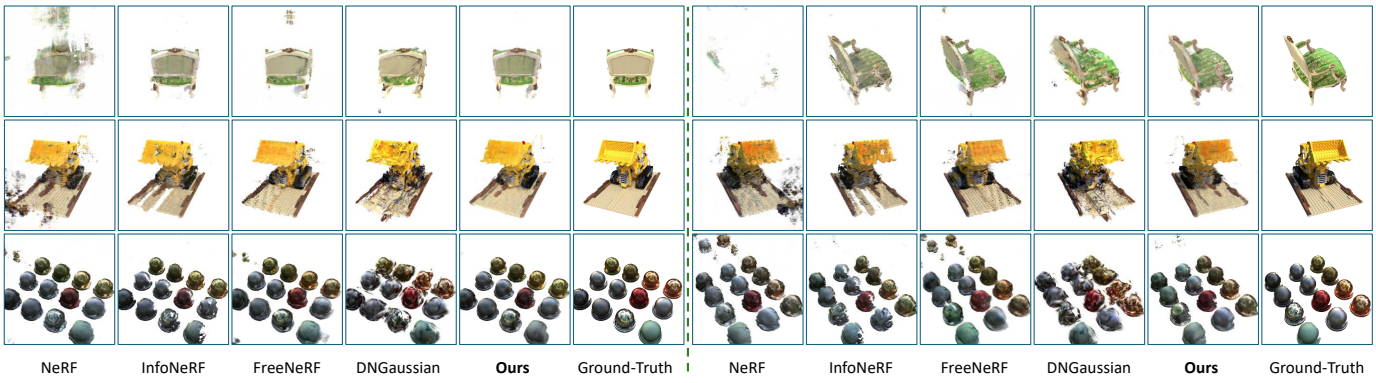


Fig. 4. Visual comparisons on selected scenes (*lego*, *chair*, *materials*) from the NeRF synthetic dataset, presented with dual perspectives. The rendering results reveal notable limitations: NeRF has noise and incomplete geometry. InfoNeRF improves geometry but keeps flaws. FreeNeRF enhances details with minor distortions. DNGaussian recovers completeness with blurry edges. In contrast, AdaGA-NeRF excels in geometry, texture, and color.

TABLE II
THE TABLE PRESENTS THE AVERAGE PSNR, SSIM, AND LPIPS VALUES FOR VARIOUS METHODS EVALUATED ON THE NERF SYNTHETIC DATASET, DEMONSTRATING THAT ADAGA-NERF OUTPERFORMS OTHER APPROACHES. PART OF THE COMPARATIVE DATA IS SOURCED FROM [35]. WE HIGHLIGHT THE BEST, SECOND-BEST, AND THIRD-BEST SCORES.

| Methods | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| NeRF, 100views | 31.01 | 0.947 | 0.081 |
| PixelNeRF* [31] | 16.09±0.78 | 0.738±0.012 | 0.390±0.030 |
| NeRF [1] | 15.93±1.06 | 0.780±0.014 | 0.320±0.049 |
| DietNeRF [10] | 16.06±1.13 | 0.793±0.019 | 0.306±0.050 |
| InfoNeRF [35] | 18.65±0.18 | 0.811±0.008 | 0.230±0.008 |
| FreeNeRF [36] | 19.03±1.52 | 0.806±0.012 | 0.181±0.016 |
| DNGaussian [44] | 18.96±1.23 | 0.798±0.014 | 0.169±0.015 |
| AdaGA-NeRF | 19.27±0.35 | 0.813±0.006 | 0.215±0.012 |

FreeNeRF at a comparable level) compared to AdaGA-NeRF ($0.215 \pm 0.012$), these improvements come with trade-offs. DNGaussian tends to produce overly smooth outputs that compromise fine details, while FreeNeRF's lower LPIPS is likely a result of aggressive regularization that suppresses high-frequency information. In contrast, AdaGA-NeRF strikes an optimal balance by preserving critical high-frequency details while maintaining overall structural coherence.

**Qualitative Analysis.** Qualitative comparisons on scenes such as lego, chair, and materials (see Fig. 4) further illustrate the advantages of AdaGA-NeRF. The original NeRF suffers from significant visual noise and geometry distortions under sparse-view conditions. InfoNeRF reduces noise via entropy-based regularization but produces incomplete reconstructions with evident geometric flaws. FreeNeRF preserves more structural details yet introduces subtle artifacts (e.g., faint noise above the chair), and DNGaussian recovers complete shapes but with noticeable ghosting and blurred boundaries, undermining texture realism.

In contrast, AdaGA-NeRF delivers significantly improved visual quality with precise geometry, clear textures, and faithful color reproduction. These enhancements are attributed to our adaptive geometric supervision and depth-consistent regularization, which dynamically resolve depth ambiguities and enforce global coherence. Additionally, our hierarchical detail refinement (HDR) module recovers high-frequency details, while the structural consistency reinforcement (SCR) module ensures global alignment across synthesized views.

**Summary.** Overall, these results validate our key contributions: the integration of adaptive geometry-aware supervision with novel architectural refinements effectively mitigates common sparse-view issues. In practical scenarios where viewpoint coverage is inherently limited, AdaGA-NeRF offers a robust and efficient solution for high-quality novel view synthesis.

TABLE III
THE TABLE PRESENTS THE AVERAGE PSNR, SSIM, AND LPIPS VALUES FOR VARIOUS METHODS ON THE NERF LLFF DATASET, HIGHLIGHTING THE SUPERIOR PERFORMANCE OF ADAGA-NERF. WE HIGHLIGHT THE BEST, SECOND-BEST, AND THIRD-BEST SCORES.

| Methods | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| NeRF [1] | 18.66 | 0.5728 | 0.2713 |
| InfoNeRF [35] | 9.25 | 0.2188 | 0.7701 |
| DietNeRF [10] | 11.84 | 0.3404 | 0.7396 |
| DDP-NeRF [29] | 19.19 | 0.5999 | 0.3821 |
| ViP-NeRF [45] | 19.57 | 0.6085 | 0.3593 |
| DNGaussian [44] | 19.56 | 0.6275 | 0.2959 |
| AdaGA-NeRF | 19.76 | 0.6295 | 0.2868 |

**2) NeRF LLFF Dataset Evaluation.** We further assess AdaGA-NeRF on the LLFF dataset [43], which features real-world, forward-facing scenes captured under sparse-view conditions and poses significant challenges such as complex lighting, non-Lambertian surfaces, and partial occlusions. As shown in Table III, our method is compared against SOTA baselines, including NeRF [1], InfoNeRF [35], DietNeRF [10], DDP-NeRF [29], ViP-NeRF [45], and DNGaussian [44].

**Quantitative Analysis.** As shown in Table III, AdaGA-NeRF achieves the highest performance, obtaining 19.76 dB in PSNR

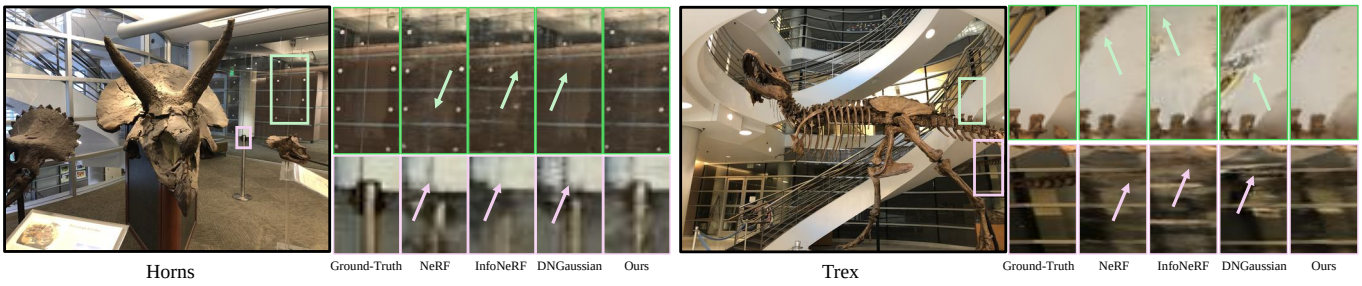


Fig. 5. Qualitative comparisons on the LLFF dataset. Two representative scenes, "*horns*" (left) and "*trex*" (right), illustrate how different methods handle challenging regions (colored boxes). NeRF suffers from blurring, InfoNeRF shows ghosting, and DNGaussian retains structure but suffers from texture artifacts, whereas AdaGA-NeRF reconstructs sharper, clearer details with robust geometric fidelity under sparse-view conditions.

and 0.6295 in SSIM, outperforming all compared methods in terms of reconstruction accuracy and structural coherence. ViP-NeRF closely follows with 19.57 dB in PSNR but slightly lags in SSIM (0.6085), highlighting AdaGA-NeRF's superior ability in maintaining structural fidelity. DNGaussian yields a competitive PSNR (19.56 dB) and the second-best SSIM (0.6275), yet it compromises fine detail preservation, as evidenced by its LPIPS (0.2959). Interestingly, the original NeRF attains the lowest LPIPS score (0.2713), but significantly underperforms in both PSNR (18.66 dB) and SSIM (0.5728), indicating locally sharp yet globally inconsistent reconstructions. In contrast, AdaGA-NeRF achieves a well-balanced LPIPS (0.2868), effectively preserving perceptual quality without sacrificing structural coherence.

**Qualitative Analysis.** Qualitative comparisons on representative scenes, "horns" and "trex" (see Fig. 5), with two highlighted regions further illustrate these differences. In the "horns" scene, NeRF exhibits noticeable blur in the regions indicated by arrows, while InfoNeRF introduces clear texture artifacts. DNGaussian further amplifies these artifacts, resulting in severe textural distortions near key structures. In contrast, AdaGA-NeRF consistently reconstructs sharp edges and coherent surfaces, aligning closely with its superior quantitative metrics (PSNR and SSIM). Similar trends are evident in the "trex" scene, where NeRF fails to capture detailed edges of the stairs, InfoNeRF produces overlapping ghosting artifacts, and DNGaussian suffers from disruptive noise patterns. Conversely, AdaGA-NeRF effectively preserves fine-grained details and global consistency, validating its adaptive scene representation and depth-consistent regularization.

**Summary.** These qualitative results strongly align with the quantitative evaluation, underscoring AdaGA-NeRF's advantage in synthesizing novel views under challenging sparse-view conditions. By adaptively modulating geometric constraints based on uncertainty, enforcing depth-consistent regularization, and leveraging architectural enhancements (HDR and SCR modules), AdaGA-NeRF robustly balances realism with structural accuracy, establishing itself as a practical and superior solution for real-world novel view synthesis tasks.

**3) NeRF Real 360 Dataset Evaluation.** We finally evaluate AdaGA-NeRF on the NeRF Real 360 dataset [1], which presents real-world scenes with significant viewpoint variation, complex geometries, reflective surfaces, and challenging lighting conditions. We compare AdaGA-NeRF with NeRF, InfoNeRF, and DNGaussian under sparse input conditions (4-view and 8-view setups).

TABLE IV
THE TABLE PRESENTS THE AVERAGE PSNR, SSIM, AND LPIPS VALUES FOR VARIOUS METHODS ON THE NERF REAL 360 DATASET, DEMONSTRATING THE SUPERIOR PERFORMANCE OF ADAGA-NERF. WE HIGHLIGHT THE BEST, SECOND-BEST, AND THIRD-BEST SCORES.

| Methods | PSNR↑ | | SSIM↑ | | LPIPS↓ | |
|---|---|---|---|---|---|---|
| | 4-view | 8-view | 4-view | 8-view | 4-view | 8-view |
| InfoNeRF | 12.07 | 17.23 | 0.2236 | 0.3596 | 0.5359 | 0.4508 |
| NeRF | 15.78 | 17.91 | 0.2921 | 0.3819 | 0.4817 | 0.4480 |
| DNGaussian [44] | 12.85 | 14.28 | 0.2512 | 0.3215 | 0.6051 | 0.5832 |
| AdaGA-NeRF | 15.86 | 17.97 | 0.2945 | 0.3854 | 0.5056 | 0.4698 |

**Quantitative Analysis.** As shown in Table IV, AdaGA-NeRF achieves the highest PSNR values under both 4-view (15.86 dB) and 8-view (17.97 dB) conditions, outperforming NeRF (15.78 dB and 17.91 dB, respectively). This superior PSNR performance underscores AdaGA-NeRF's ability to reconstruct images with greater pixel-level fidelity. Additionally, our method attains the highest SSIM scores (0.2945 for 4-view and 0.3854 for 8-view setups), highlighting its capability to maintain structural integrity and fine detail preservation in complex scenarios. While NeRF achieves better LPIPS scores (0.4817 and 0.4480) due to its inherent tendency to generate locally sharper textures, these advantages come at the expense of overall coherence and geometric accuracy. InfoNeRF and DNGaussian lag significantly behind in PSNR and SSIM, with DNGaussian performing notably worse in perceptual quality (LPIPS scores of 0.6051 and 0.5832).

**Qualitative Analysis.** Fig. 6 presents qualitative comparisons using the "vasedeck" and "pinecone" scenes. InfoNeRF suffers from severe blurring and ghosting, particularly noticeable at object boundaries, leading to ambiguity and blending of foreground and background elements. DNGaussian shows significant structural confusion, with foreground objects incorrectly bleeding into the background, resulting in prominent artifacts. In contrast, AdaGA-NeRF consistently reconstructs clearer and sharper object boundaries, accurately capturing both detailed textures and overall scene geometry (e.g., clearer background textures like the wood board in the depth map compared to NeRF). These visual results validate AdaGA-NeRF's effectiveness in adaptively focusing supervision on uncertain regions and enforcing depth consistency, substan-

TABLE V
PERFORMANCE COMPARISON OF ADAGA-NERF MODULES ON THE NERF LLFF DATASET

| Method | $\mathcal{L}_{\text{adaptive}}$ | $\mathcal{L}_{\text{depth}}$ | $\mathcal{M}_{\text{HDR}}$ | $\mathcal{M}_{\text{SCR}}$ | PSNR ↑ | SSIM ↑ | LPIPS ↓ |
|---|---|---|---|---|---|---|---|
| Baseline | | | | | 9.25 | 0.2188 | 0.7701 |
| AdaGA-NeRF | ✓ | ✗ | ✗ | ✗ | 18.21 | 0.4823 | 0.3964 |
| AdaGA-NeRF | ✓ | ✓ | ✗ | ✗ | 18.93 | 0.5515 | 0.3130 |
| AdaGA-NeRF | ✓ | ✓ | ✓ | ✗ | 19.17 | 0.5391 | 0.3139 |
| AdaGA-NeRF | ✓ | ✓ | ✓ | ✓ | 19.76 | 0.6295 | 0.2868 |

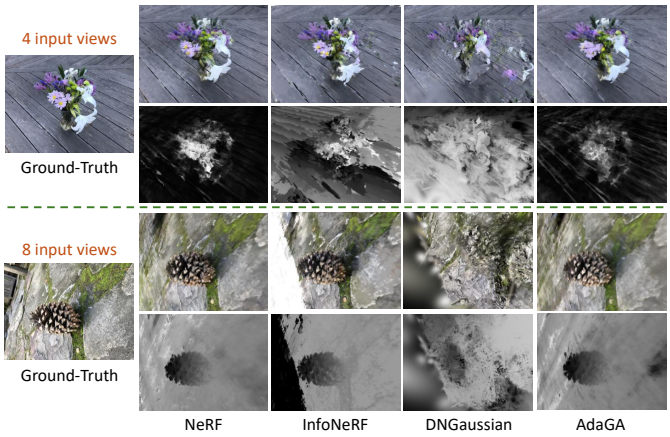


Fig. 6. Qualitative comparisons on the real 360 dataset ( scenes "*vasedeck*" (top) and "*pinecone*" (bottom) ). Compared with other methods, our method achieves clearer textures, sharper boundaries, and superior structural consistency.

tially mitigating artifacts common in sparse-view scenarios.

**Summary.** Overall, the quantitative and qualitative results highlight AdaGA-NeRF's robust capability to address the challenges posed by real-world 360-degree scenes. By leveraging adaptive geometric constraints and architectural enhancements (HDR and SCR modules), AdaGA-NeRF achieves superior reconstruction fidelity, demonstrating broad applicability and effectiveness in practical, sparse-view novel view synthesis tasks.

### C. Ablation Study

**1) Module Effectiveness Analysis.** We conduct an ablation study on the LLFF dataset to quantify the effectiveness of each proposed AdaGA-NeRF module (Table V and Fig. 7). Starting from a baseline InfoNeRF (PSNR: 9.25, SSIM: 0.2188, LPIPS: 0.7701), incorporating the adaptive geometric supervision term $\mathcal{L}_{\text{adaptive}}$ markedly enhances reconstruction quality (PSNR: 18.21, SSIM: 0.4823, LPIPS: 0.3964), effectively addressing sparse-view ambiguities by dynamically focusing on uncertain regions. Further integrating the depth-consistent regularization $\mathcal{L}_{\text{depth}}$ improves global geometric coherence significantly (PSNR: 18.93, SSIM: 0.5515, LPIPS: 0.3130), clearly mitigating depth inconsistencies across neighboring rays. The addition of the Hierarchical Detail Refinement module ($\mathcal{M}_{\text{HDR}}$) further sharpens local details and enhances texture clarity (PSNR: 19.17, SSIM: 0.5391, LPIPS: 0.3139). Finally, with the Structural Consistency Reinforcement module ($\mathcal{M}_{\text{SCR}}$), the complete AdaGA-NeRF achieves superior performance (PSNR: 19.76, SSIM: 0.6295, LPIPS: 0.2868), balancing detailed local representation with global structural integrity.

Qualitative visualizations reinforce these quantitative observations (Fig. 7). The baseline produces blurred textures and distorted depth maps (e.g., ceiling tiles and telephone structure in the room scene). Introducing $\mathcal{L}_{\text{adaptive}}$ notably reduces artifacts but reveals persistent depth gaps. Adding $\mathcal{L}_{\text{depth}}$ further enhances geometry coherence, improving clarity of major structures. Integrating $\mathcal{M}_{\text{HDR}}$ visibly sharpens finer local details, though minor inconsistencies remain. The final AdaGA-NeRF model, augmented by $\mathcal{M}_{\text{SCR}}$, achieves crisp, realistic renderings, demonstrating the cumulative effectiveness of our designed modules in overcoming sparse-view challenges.

**2) Sensitivity Analysis of Hyperparameters $\beta$ and $\lambda$.** To assess the influence of hyperparameters $\beta$ (entropy weighting steepness) and $\lambda$ (depth-adaptive regularization decay), we conducted sensitivity experiments on the LLFF dataset under sparse-view conditions. We varied $\beta \in \{1, 5, 10\}$ and $\lambda \in \{0.1, 5, 10\}$, evaluating reconstruction quality via PSNR, SSIM, and LPIPS metrics (Fig. 8).

The optimal hyperparameter configuration ($\beta = 1, \lambda = 0.1$) achieves the highest PSNR (20.86), competitive SSIM (0.4830), and lowest LPIPS (0.2755). Lower values of $\beta$ produce gentler weighting curves, allowing balanced supervision across rays with varied depth uncertainty, thus stabilizing training and effectively capturing complex geometries. Conversely, larger values of $\beta$ excessively concentrate supervision, potentially destabilizing the learning process in noisy regions. Similarly, a smaller $\lambda$ (e.g., 0.1) maintains global coherence by gently decaying depth-based regularization, whereas higher values (e.g., 10) overly penalize outliers, harming reconstruction fidelity.

These insights highlight the necessity of balanced hyperparameter selection to achieve robust and coherent scene reconstructions. Our optimal parameter setting confirms the designed trade-off, leveraging adaptive entropy-based supervision to address local uncertainties, and modest depth-adaptive regularization to enforce global structural consistency, ultimately resulting in superior sparse-view performance.

**3) Overfitting Evaluation.** Limited input views inherently raise the risk of model overfitting, adversely affecting generalization to unseen perspectives. To evaluate the robustness of AdaGA-NeRF against overfitting, we perform comparative experiments with InfoNeRF using the "pinecone" scene from

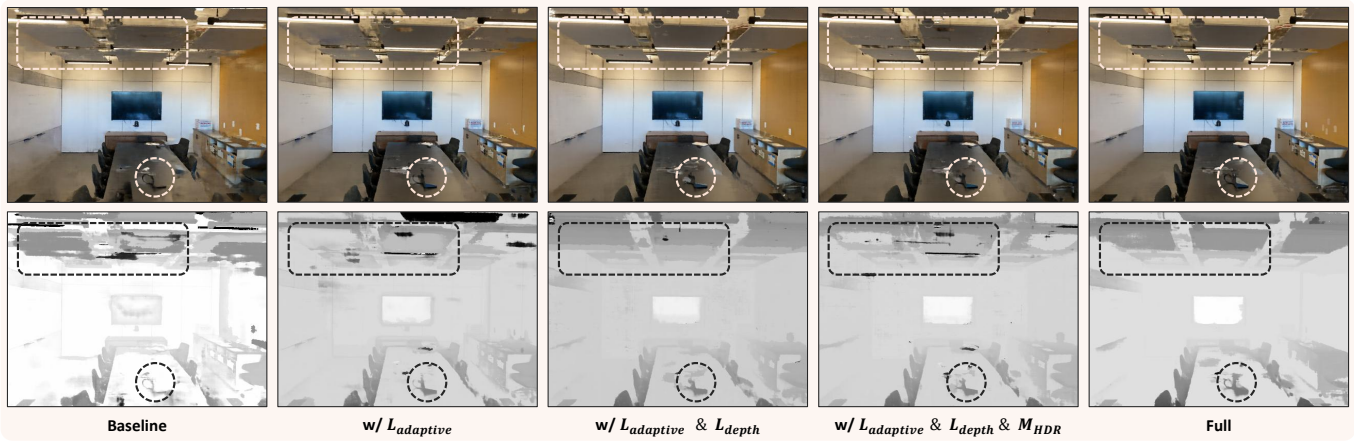


Fig. 7. **Visual ablation study on the LLFF *room* scene.** We compare the baseline (InfoNeRF) with AdaGA-NeRF configurations incrementally adding adaptive supervision ($\mathcal{L}_{\text{adaptive}}$), depth-consistent regularization ($\mathcal{L}_{\text{depth}}$), hierarchical detail refinement ($\mathcal{M}_{\text{HDR}}$), and structural consistency reinforcement ($\mathcal{M}_{\text{SCR}}$). The baseline shows significant noise and structural distortion. Adding $\mathcal{L}_{\text{adaptive}}$ reduces noise, $\mathcal{L}_{\text{depth}}$ improves global consistency, and $\mathcal{M}_{\text{HDR}}$ enhances local details. The full model achieves accurate geometry and realistic appearance, highlighting the cumulative effect of the proposed modules.

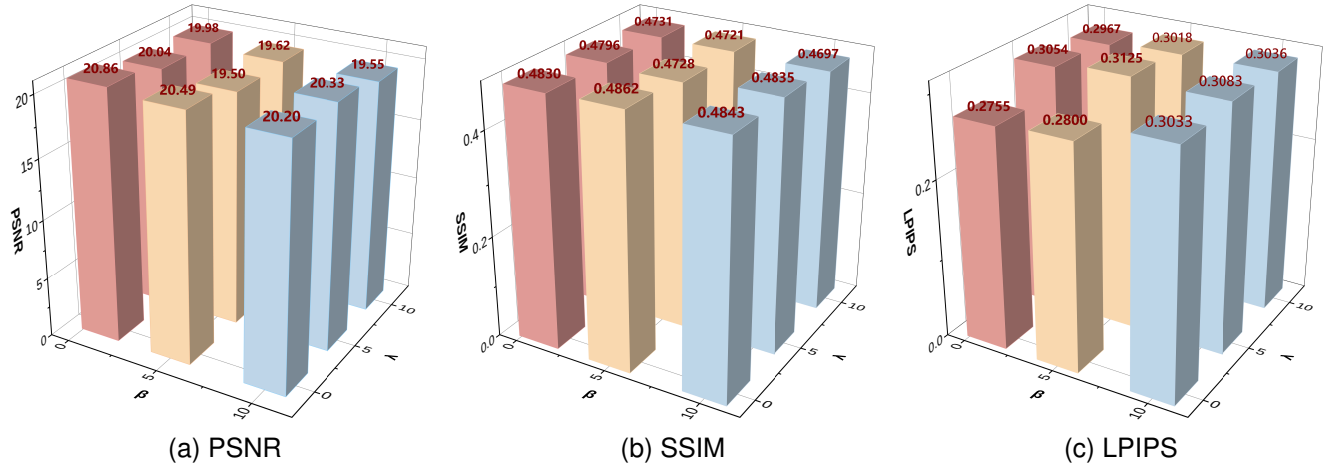


Fig. 8. Three-dimensional bar plots evaluating the impact of hyperparameters $\beta$ and $\lambda$ on model reconstruction quality. Subfigures (a), (b), and (c) represent PSNR, SSIM, and LPIPS metrics, respectively.

the NeRF Real 360 dataset under sparse-view settings (4-view and 8-view). As illustrated in Fig. 9, when trained with only 4 input views, both AdaGA-NeRF and InfoNeRF achieve similar training performance initially. However, AdaGA-NeRF significantly outperforms InfoNeRF on the test set, highlighting InfoNeRF's greater susceptibility to overfitting. When the number of input views increases to 8, InfoNeRF's overfitting issue is partially alleviated, yet its test performance still notably trails behind our approach. These results confirm AdaGA-NeRF's superior ability to resist overfitting and maintain strong generalization, particularly under challenging sparse-view conditions.

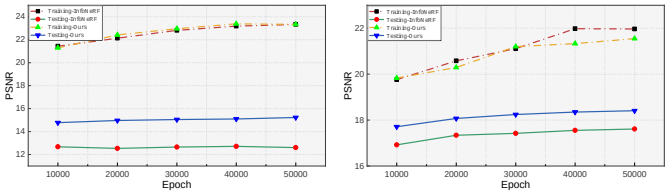


Fig. 9. PSNR comparison on the "*pinecone*" scene with 4 views (left) and 8 views (right). AdaGA-NeRF achieves comparable training PSNR but significantly higher testing PSNR, indicating stronger generalization and reduced overfitting under sparse views.

### *D. Limitations*

While our proposed AdaGA-NeRF framework has demonstrated significant improvements over existing sparse-view novel view synthesis methods, we acknowledge several areas that may be further explored. One notable aspect is the sensitivity of our adaptive geometric supervision to hyperparameters, specifically $\beta$ and $\lambda$. Although our experiments have identified effective empirical settings, systematically determining these parameters across diverse scenarios may require additional investigation. This challenge naturally leads to questions of generalization, as AdaGA-NeRF achieves robust performance across standard sparse-view datasets, yet its applicability to extreme sparsity scenarios, such as those with fewer than four views or highly non-standard viewpoints, warrants

further exploration. Assessing performance under these highly constrained conditions remains a valuable direction for future work. Moreover, the introduction of architectural refinements, such as the HDR and SCR modules, while enhancing quality, results in modest computational overhead. Although this overhead is manageable for typical scenarios, optimizing the approach for real-time inference or deployment on resource-limited devices may necessitate further enhancements in model efficiency. These observations do not detract from the overall contributions and significance of AdaGA-NeRF but instead highlight opportunities for ongoing refinement and extension.

## V. Conclusion

In this paper, we introduced AdaGA-NeRF, a fully self-contained NeRF framework specifically designed to tackle the sparse-view challenge without reliance on external priors. Our method integrates adaptive geometry-aware supervision, depth-consistent regularization, hierarchical detail refinement (HDR), and structural consistency reinforcement (SCR), collectively addressing geometric ambiguity, texture blurring, and structural inconsistencies commonly encountered in sparse-view novel view synthesis.

Extensive evaluations across the NeRF Synthetic, LLFF, and Real 360 datasets demonstrate that AdaGA-NeRF consistently outperforms state-of-the-art methods in both quantitative metrics (PSNR, SSIM, LPIPS) and qualitative visual fidelity. Ablation studies validated each module's individual and collective contributions, highlighting their synergistic effects on reconstruction quality. Furthermore, our sensitivity analysis of hyperparameters ($\beta$ and $\lambda$) underscored the necessity of balanced geometric supervision for achieving robust performance.

Overall, AdaGA-NeRF provides a practical, effective, and computationally reasonable solution for sparse-view novel view synthesis. Beyond the evaluated benchmarks, its robustness to limited viewpoints and ability to preserve fine details make it well-suited for practical scenarios such as VR-based 3D scene reconstruction, where capturing dense viewpoints is often infeasible, and theatrical stage digitization, which demands structural fidelity under complex lighting. Future research will further explore theoretical analyses of adaptive weighting strategies, improve computational efficiency, and extend our approach to address even more diverse and challenging real-world scenarios, including those with extreme sparsity or dynamic scene changes.

## References


[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, p. 99–106, Dec. 2021.

[2] N. Deng, Z. He, J. Ye, B. Duinkharjav, P. Chakravarthula, X. Yang, and Q. Sun, "Fov-nerf: Foveated neural radiance fields for virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3854–3864, 2022.

[3] C. Li, S. Li, Y. Zhao, W. Zhu, and Y. Lin, "Rt-nerf: Real-time on-device neural radiance fields towards immersive ar/vr rendering," in *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design*, ICCAD '22, (New York, NY, USA), Association for Computing Machinery, 2022.

[4] M. Adamkiewicz, T. Chen, A. Caccavale, R. Gardner, P. Culbertson, J. Bohg, and M. Schwager, "Vision-only robot navigation in a neural radiance world," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4606–4613, 2022.

[5] O. Kwon, J. Park, and S. Oh, "Renderable neural radiance map for visual navigation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9099–9108, June 2023.

[6] C.-H. Lin, J. Gao, L. Tang, T. Takikawa, X. Zeng, X. Huang, K. Kreis, S. Fidler, M.-Y. Liu, and T.-Y. Lin, "Magic3d: High-resolution text-to-3d content creation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 300–309, June 2023.

[7] Y. Chen, J. Zhang, Z. Xie, W. Li, F. Zhang, J. Lu, and L. Zhang, "S-nerf++: Autonomous driving simulation via neural reconstruction and generation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 6, pp. 4358–4376, 2025.

[8] G. Wang, Z. Chen, C. C. Loy, and Z. Liu, "Sparsenerf: Distilling depth ranking for few-shot novel view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9065–9076, October 2023.

[9] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, "Depth-supervised nerf: Fewer views and faster training for free," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12882–12891, June 2022.

[10] A. Jain, M. Tancik, and P. Abbeel, "Putting nerf on a diet: Semantically consistent few-shot view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5885–5894, October 2021.

[11] C. Wang, M. Chai, M. He, D. Chen, and J. Liao, "Clip-nerf: Text-and-image driven manipulation of neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3835–3844, June 2022.

[12] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, "Barf: Bundle-adjusting neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5741–5751, October 2021.

[13] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "Nerf in the wild: Neural radiance fields for unconstrained photo collections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7210–7219, June 2021.

[14] P. P. Srinivasan, B. Deng, X. Zhang, M. Tancik, B. Mildenhall, and J. T. Barron, "Nerv: Neural reflectance and visibility fields for relighting and view synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7495–7504, June 2021.

[15] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds.), vol. 33, pp. 7537–7547, Curran Associates, Inc., 2020.

[16] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. M. Sajjadi, A. Geiger, and N. Radwan, "Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5480–5490, June 2022.

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

[18] J. Tang, H. Zhou, X. Chen, T. Hu, E. Ding, J. Wang, and G. Zeng, "Delicate textured mesh recovery from nerf via adaptive surface refinement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17739–17749, October 2023.

[19] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5855–5864, October 2021.

[20] L. Yen-Chen, P. Florence, J. T. Barron, T.-Y. Lin, A. Rodriguez, and P. Isola, "Nerf-supervision: Learning dense object descriptors from neural radiance fields," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 6496–6503, 2022.

[21] F. Wan, X. Miao, H. Duan, J. Deng, R. Gao, and Y. Long, "Sentinel-guided zero-shot learning: A collaborative paradigm without real data exposure," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 9, pp. 8067–8079, 2024.

[22] F. Wan, J. Wang, H. Duan, Y. Song, M. Pagnucco, and Y. Long, "Community-aware federated video summarization," in *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2023.

[23] J. Wang, Z. Sun, Z. Tan, X. Chen, W. Chen, H. Li, C. Zhang, and Y. Song, "Towards effective usage of human-centric priors in diffusion models for text-based human image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8446–8455, June 2024.
[24] M. Shao, X. Miao, H. Duan, Z. Wang, J. Chen, Y. Huang, X. Wu, J. Deng, Y. Long, and Y. Zheng, "Trace: Temporally reliable anatomically-conditioned 3d ct generation with enhanced efficiency," 2025.
[25] X. Miao, Y. Bai, H. Duan, F. Wan, Y. Huang, Y. Long, and Y. Zheng, "Conrf: Zero-shot stylization of 3d scenes with conditioned radiation fields," 2024.
[26] X. Miao, Y. Bai, H. Duan, F. Wan, Y. Huang, Y. Long, and Y. Zheng, "Ctnerf: Cross-time transformer for dynamic neural radiance field from monocular video," *Pattern Recognition*, vol. 156, p. 110729, 2024.
[27] T. Zhang, F. Wan, H. Duan, K. W. Tong, J. Deng, and Y. Long, "Fmdconv: Fast multi-attention dynamic convolution via speed-accuracy trade-off," *Knowledge-Based Systems*, vol. 317, p. 113393, 2025.
[28] Y. Li, F. Wan, and Y. Long, "Sid-nerf: Few-shot nerf based on scene information distribution," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2024.
[29] B. Roessle, J. T. Barron, B. Mildenhall, P. P. Srinivasan, and M. Nießner, "Dense depth priors for neural radiance fields from sparse input views," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12892–12901, June 2022.
[30] Y.-J. Yuan, Y.-K. Lai, Y.-H. Huang, L. Kobbelt, and L. Gao, "Neural radiance fields from sparse rgb-d images for high-quality view synthesis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8713–8728, 2023.
[31] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelnerf: Neural radiance fields from one or few images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4578–4587, June 2021.
[32] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14124–14133, October 2021.
[33] S. Kobayashi, E. Matsumoto, and V. Sitzmann, "Decomposing nerf for editing via feature field distillation," in *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 23311–23330, Curran Associates, Inc., 2022.
[34] R. He, S. Huang, X. Nie, T. Hui, L. Liu, J. Dai, J. Han, G. Li, and S. Liu, "Customize your nerf: Adaptive source driven 3d scene editing via local-global iterative training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6966–6975, June 2024.
[35] M. Kim, S. Seo, and B. Han, "Infonerf: Ray entropy minimization for few-shot neural volume rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12912–12921, June 2022.
[36] J. Yang, M. Pavone, and Y. Wang, "Freenerf: Improving few-shot neural rendering with free frequency regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8254–8263, June 2023.
[37] Y. Yu, R. Wu, Y. Men, S. Lu, M. Cui, X. Xie, and C. Miao, "Morphnerf: Text-guided 3d-aware editing via morphing generative neural radiance fields," *IEEE Transactions on Multimedia*, vol. 26, pp. 8516–8528, 2024.
[38] Z. Luo, A. Rocha, B. Shi, Q. Guo, H. Li, and R. Wan, "The nerf signature: Codebook-aided watermarking for neural radiance fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 6, pp. 4652–4667, 2025.
[39] Y. Chen, L. Zhang, S. Zhao, and Y. Zhou, "Atm-nerf: Accelerating training for nerf rendering on mobile devices via geometric regularization," *IEEE Transactions on Multimedia*, vol. 27, pp. 3279–3293, 2025.
[40] M.-S. Kwak, J. Song, and S. Kim, "GeCoNeRF: Few-shot neural radiance fields via geometric consistency," in *Proceedings of the 40th International Conference on Machine Learning* (A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, eds.), vol. 202 of *Proceedings of Machine Learning Research*, pp. 18023–18036, PMLR, 23–29 Jul 2023.
[41] H. Fu, X. Yu, L. Li, and L. Zhang, "Cbarf: Cascaded bundle-adjusting neural radiance fields from imperfect camera poses," *IEEE Transactions on Multimedia*, vol. 26, pp. 9304–9315, 2024.
[42] B. Liu, J. Lei, B. Peng, Z. Zhang, J. Zhu, and Q. Huang, "Advancing generalizable occlusion modeling for neural human radiance field," *IEEE Transactions on Multimedia*, vol. 27, pp. 1362–1373, 2025.
[43] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: practical view synthesis with prescriptive sampling guidelines," *ACM Trans. Graph.*, vol. 38, July 2019.
[44] J. Li, J. Zhang, X. Bai, J. Zheng, X. Ning, J. Zhou, and L. Gu, "Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20775–20785, June 2024.
[45] N. Somraj and R. Soundararajan, "Vip-nerf: Visibility prior for sparse input neural radiance fields," in *ACM SIGGRAPH 2023 Conference Proceedings*, SIGGRAPH '23, (New York, NY, USA), Association for Computing Machinery, 2023.

**Fan Wan** received the M.Sc. degree (with Distinction) in Computer Science from Newcastle University, UK, in 2018, and the Ph.D. degree in Computer Science from Durham University, UK, in January 2025. He is currently a Researcher with Tongfang Knowledge Network Digital Technology Co., Ltd., China National Nuclear Corporation, Beijing, China. His research focuses on the application of large language models (LLMs) in nuclear industry scenarios. His interests include machine learning, computer vision, multimedia analysis, and the development and application of LLMs, including LLM-based agents and downstream tasks.

**Zhihan Jiang** holds a bachelor's degree in Theatre Studies from the Central Academy of Drama and is currently pursuing a master's degree in Theatre and Performance Studies at the National University of Singapore. Her research interests include immersive theatre, digital theatre, and the integration of cultural tourism with performance. She is currently focusing on institutional and policy innovations that facilitate the deep integration of digital technology and theatre arts, with a particular interest in applying NeRF and other AI techniques to enhance virtual scenography in digital performances.

**Yuchen Li** received his M.Sc. degree in Electronics from Queen's University Belfast, UK, in 2022. He is a PhD student in the Department of Computer Science, Durham University, UK. His research interests focus on novel view synthesis and related areas.

**Xingyu Miao** received a master's degree from the School of Information Engineering, Ningxia University, China. Currently, he is a PhD student in the Department of Computing, Durham University, UK. His current main focus is monocular depth estimation and 3D reconstruction.

**Xueqi Qiu** received the M.Sc. degree in Advanced Computer Science from Newcastle University, U.K., in 2022. She is currently pursuing a Ph.D. degree with the Department of Computer Science, Durham University. Her research interests include computer vision, machine learning and 3D reconstruction.

**Tianyu Zhang** received a Distinction M.S. degree from Northeastern University USA in 2020. Currently, he is a PhD student in the Department of Computing, at Durham University, UK. His current main focuses are dynamic convolution and video out-painting.

**Yang Long** is an Associate Professor in the Department of Computer Science, Durham University. He is also an IEEE Senior Member (SMIEEE) and MRC Innovation Fellow, aiming to design scalable AI solutions for large-scale healthcare applications. His research background is in the highly interdisciplinary field of Computer Vision and Machine Learning. While he is passionate about unveiling the black-box of AI brain and transferring the knowledge to seek Scalable, interactable, interpretable, and sustainable solutions for other disciplinary research, e.g., physical activity, mental health, design, education, security, and geo-engineering. He has authored/co-authored 100+ top-tier papers in refereed journals/conferences such as IEEE TPAMI, TIP, CVPR, AAAI, and ACM MM.